ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ

Διπλωματική Εργασία

# ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ

της

ΓΟΥΛΟΥ ΖΩΗΣ του ΑΣΤΕΡΙΟΥ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΓΕΩΡΓΙΟΥ ΑΝΔΡΕΑΣ

Υποβλήθηκε ως απαιτούμενο για την απόκτηση του μεταπτυχιακού διπλώματος ειδίκευσης στη Διοίκηση Επιχειρήσεων (με εξειδίκευση στο Μάρκετινγκ)

Φεβρουάριος 2012

**ΑΦΙΕΡΩΣΕΙΣ**

Η διπλωματική αυτή εργασία
αφιερώνεται στους γονείς μου,
Αστέριο και Λεμονιά.

**ΕΥΧΑΡΙΣΤΙΕΣ**

Θα ήθελα να ευχαριστήσω θερμά τον κ. Ανδρέα Γεωργίου, πρόεδρο του Δ.Π.Μ.Σ. στη Διοίκηση Επιχειρήσεων του Πανεπιστημίου Μακεδονίας, και τον κ. Ανδρέα Ανδρονικίδη, επίκουρο καθηγητή του τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων του Πανεπιστημίου Μακεδονίας, για την πολύτιμη βοήθεια και καθοδήγηση που μου προσέφεραν καθ' όλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας.

**ΠΕΡΙΛΗΨΗ**

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι να εξετάσει βιβλιογραφικά την εφαρμογή των μεθόδων εξόρυξης δεδομένων στις στρατηγικές διαχείρισης πελατειακών σχέσεων. Αρχικά γίνεται μελέτη της έννοιας, των βασικών αλγορίθμων και των βημάτων διαδικασίας της εξόρυξης δεδομένων και αναφέρονται τα είδη πληροφοριών που εξάγονται από την εφαρμογή των αλγορίθμων. Κατόπιν, αναλύονται οι έννοιες και τα στάδια ανάπτυξης ενός συστήματος διαχείρισης πελατειακών σχέσεων, εξετάζονται οι τέσσερις διαστάσεις του CRM και αναφέρονται οι δύο κύριες λειτουργίες του.

Στη συνέχεια, γίνεται η ταξινόμηση των τεχνικών εξόρυξης δεδομένων, ανάλογα με τη διάσταση του CRM πάνω στην οποία εφαρμόζονται, παραθέτονται τα βήματα για αποδοτικές CRM εφαρμογές, τονίζεται η σημαντικότητα της εξόρυξης δεδομένων στον κύκλο ζωής του πελάτη και γίνεται θεωρητική ανάλυση των λειτουργιών του CRM με τη χρήση εργαλείων της εξόρυξης δεδομένων. Η εργασία συνεχίζει με την εξέταση του όρου web mining, το οποίο αποτελεί την εφαρμογή των τεχνικών εξόρυξης δεδομένων στα δεδομένα του διαδικτύου. Στην ίδια ενότητα, αναφέρονται οι κατηγορίες και η δομή του, αναλύονται τα εργαλεία ανακάλυψης προτύπων που χρησιμοποιεί και γίνεται ιδιαίτερη αναφορά στην εφαρμογή του στο ηλεκτρονικό εμπόριο και τη διαχείριση πελατειακών σχέσεων.

Η εξέταση περιπτώσεων εφαρμογής των μεθόδων εξόρυξης δεδομένων στη διαχείριση πελατειακών σχέσεων οργανισμών διαφορετικών κλάδων κρίθηκε αναγκαία, για την βαθύτερη κατανόηση του γνωστικού αντικειμένου. Για το λόγο αυτό μελετήθηκαν οι εφαρμογές της εξόρυξης στην τουριστική βιομηχανία και τον κλάδο της υγείας. Η μελέτη ολοκληρώνεται με την παρουσίαση των ηθικών προεκτάσεων της εξόρυξης δεδομένων, ένα θέμα αυξανόμενης ανησυχίας των ατόμων για την προστασία των προσωπικών τους δεδομένων.

**Λέξεις – κλειδιά**: Εξόρυξη Δεδομένων, Διαχείριση Πελατειακών Σχέσεων, CRM, Web Mining, Ηθική.

**ΚΑΤΑΛΟΓΟΣ ΠΕΡΙΕΧΟΜΕΝΩΝ**

ΑΦΙΕΡΩΣΕΙΣ.....ii
ΕΥΧΑΡΙΣΤΙΕΣ.....iii
ΠΕΡΙΛΗΨΗ.....iv
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ.....v
ΠΙΝΑΚΑΣ ΕΙΚΟΝΟΓΡΑΦΗΣΕΩΝ.....viii

**ΚΕΦΑΛΑΙΟ 1: Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ**.....1
1.1 Εισαγωγή.....1
1.2 Έννοιες – Ορισμοι.....1
1.3 Τα βήματα της εξόρυξης δεδομένων.....3
1.4 Οι βασικοί αλγόριθμοι της εξόρυξης δεδομένων.....4
1.5 Ο σκοπός και τα μοντέλα πληροφοριών της εξόρυξης δεδομένων.....4
1.6 Ο βασικός διαχωρισμός της εξόρυξης δεδομένων.....8
1.6.1 Η κατευθυνόμενη εξόρυξη δεδομένων.....8
1.6.2 Η μη-κατευθυνόμενη εξόρυξη δεδομένων.....8

**ΚΕΦΑΛΑΙΟ 2: Η ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ (CRM)**.....9
2.1 Εισαγωγή.....9
2.2 Έννοιες – Ορισμοί.....9
2.3 Τα στάδια ανάπτυξης ενός CRM συστήματος.....11
2.4 Οι τέσσερις διαστάσεις του CRM.....13
2.5 Συλλογή των δεδομένων των πελατών.....14
2.6 Τμηματοποίηση πελατών και δημιουργία προφίλ πελατών.....16
2.6.1 Τμηματοποίηση πελατών.....17
2.6.2 Δημιουργία προφίλ πελατών.....18
2.7 Επιχειρησιακό CRM και Αναλυτικό CRM.....20

**ΚΕΦΑΛΑΙΟ 3: Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ**.....22
3.1 Εισαγωγή.....22
3.2 Έννοιες – Ορισμοί .....23
3.3 Ταξινόμηση των τεχνικών εξόρυξης δεδομένων στις τέσσερις διαστάσεις του CRM.....23

3.4 Προετοιμασία των δεδομένων για τη διαδικασία της εξόρυξης δεδομένων.....25
3.4.1 Προετοιμασία των δεδομένων για όλα τα εργαλεία εξόρυξης δεδομένων.....25
3.4.2 Προετοιμασία των δεδομένων ανάλογα με το χρησιμοποιούμενο εργαλείο εξόρυξης δεδομένων.....27
3.5 Τα βήματα εφαρμογής της εξόρυξης δεδομένων στη διαχείριση πελατειακών σχέσεων.....28
3.6 Η εξόρυξη δεδομένων στον κύκλο ζωής του πελάτη.....30
3.7 Marketing Data Intelligence.....31
3.8 Τμηματοποίηση πελατών με εργαλεία της εξόρυξης δεδομένων.....32
3.8.1 Μοντέλο τμηματοποίησης πελατών.....33
3.8.2 Πλεονεκτήματα της μεθόδου τμηματοποίησης με εργαλεία της εξόρυξης δεδομένων.....34
3.9 Δημιουργία προφίλ πελατών με εργαλεία της εξόρυξης δεδομένων.....35
3.10 Αξιολόγηση της αφοσίωσης των πελατών.....37
3.11 Ανίχνευση ακραίων τιμών (outlier detection).....39

**ΚΕΦΑΛΑΙΟ 4: ΤΟ WEB MINING ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ**.....40
4.1 Εισαγωγή.....40
4.2 Έννοιες – Ορισμοί.....41
4.3 Οι κατηγορίες του Web Mining.....41
4.3.1 Web Content Mining.....42
4.3.2 Web Usage Mining.....43
4.3.3 Web Structure Mining.....43
4.4 Η δομή του Web Mining.....43
4.5 Πηγές διαδικτυακών δεδομένων (web data).....45
4.6 Εργαλεία ανακάλυψης προτύπων-μοτίβων του web mining.....46
4.7 Το Web Mining στο ηλεκτρονικό εμπόριο .....47
4.8 Το Web Mining στη διαχείριση πελατειακών σχέσεων.....48
4.9 Ανακάλυψη του προφίλ πελατών μέσω του Web Usage Mining.....51

**ΚΕΦΑΛΑΙΟ 5: ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ**.....52
5.1 Εισαγωγή.....52
5.2 Εφαρμογή μεθόδων εξόρυξης δεδομένων στην τουριστική βιομηχανία.....53
5.2.1 Εισαγωγή.....53
5.2.2 Η εξόρυξη δεδομένων στο διαδικτυακό τουριστικό μάρκετινγκ.....54
5.2.3 Η διαδικασία της εξόρυξης δεδομένων στον τουριστικό οργανισμό.....55
5.3 Εφαρμογή μεθόδων εξόρυξης δεδομένων στη βιομηχανία της υγείας.....57
5.3.1 Εισαγωγή.....57
5.3.2 Εφαρμογή της εξόρυξης δεδομένων στη διαχείριση πελατειακών σχέσεων της νοσοκομειακής μονάδας.....58
5.3.3 Άλλες κατηγορίες εφαρμογής μεθόδων εξόρυξης δεδομένων στο νοσοκομειακό επίπεδο.....61
5.4 Εφαρμογή της εξόρυξης δεδομένων σε άλλους κλάδους.....62
5.4.1 Εφαρμογή της εξόρυξης δεδομένων στον τραπεζικό τομέα.....62
5.4.2 Εφαρμογή της εξόρυξης δεδομένων στο λιανικό εμπόριο.....63
5.4.3 Εφαρμογή της εξόρυξης δεδομένων στον κλάδο τηλεπικοινωνιών.....63

**ΚΕΦΑΛΑΙΟ 6: ΗΘΙΚΑ ΖΗΤΗΜΑΤΑ ΣΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ**.....64
6.1 Εισαγωγή.....64
6.2 Έννοιες – Ορισμοί.....65
6.3 Προστασία των προσωπικών δεδομένων.....65
6.4 Ακρίβεια των δεδομένων.....67
6.5 Ασφάλεια της βάσης δεδομένων.....67
6.6 Εξόρυξη δεδομένων & παραδοσιακές τεχνικές ανάκτησης πληροφοριών.....68
6.7 Νέες απειλές των προσωπικών δεδομένων.....69
6.8 Τεχνικές για την προστασία των προσωπικών δεδομένων στην εξόρυξη δεδομένων.....70

**ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ**.....72
7.1 Συμπεράσματα.....72
7.2 Περιορισμοί και προτάσεις για περεταίρω έρευνα.....76

**ΠΑΡΑΡΤΗΜΑ**.....78
**ΒΙΒΛΙΟΓΡΑΦΙΑ**.....84

**ΠΙΝΑΚΑΣ ΕΙΚΟΝΟΓΡΑΦΗΣΕΩΝ**

**ΚΕΦΑΛΑΙΟ 1: Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ**

Εικόνα 1.1: Η διαδικασία της εξόρυξης δεδομένων.....3

Εικόνα 1.2: Ταξινόμηση των μοντέλων πληροφοριών της εξόρυξης δεδομένων.....7

**ΚΕΦΑΛΑΙΟ 2: Η ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ (CRM)**

Εικόνα 2.1: Οι τέσσερις διαστάσεις του CRM.....13

Εικόνα 2.2: Η CRM διαδικασία με βάση τον κύκλο ζωής του πελάτη.....21

Εικόνα 2.3: Ένταξη των στόχων του CRM στους στόχους της επιχείρησης.....22

**ΚΕΦΑΛΑΙΟ 3: Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ**

Εικόνα 3.1: Ταξινόμηση των τεχνικών εξόρυξης δεδομένων στο CRM.....24

Εικόνα 3.2: Μοντέλο τμηματοποίησης πελατών με βάση την εξόρυξη δεδομένων.....34

Εικόνα 3.3: Σύστημα δημιουργίας προφίλ πελατών.....36

Εικόνα 3.4: Μέθοδος αξιολόγησης της αφοσίωσης των πελατών.....38

**ΚΕΦΑΛΑΙΟ 4: Η ΔΙΑΔΙΚΤΥΑΚΗ ΕΞΟΡΥΞΗ (WEB MINING) ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ**

Εικόνα 4.1: Οι κατηγορίες του web mining.....42

Εικόνα 4.2 Η δομή του Webminer.....44

Εικόνα 4.3: Μοντέλο web mining στο CRM.....50

Εικόνα 4.4: Διαδικασία του Web Usage Mining για την ανακάλυψη του προφίλ των χρηστών.....52

**ΚΕΦΑΛΑΙΟ 5: ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ**

Εικόνα 5.1: Διαδικασία εξόρυξης δεδομένων στον ταξιδιωτικό οργανισμό.....56

Εικόνα 5.2: Επίπεδα εφαρμογής της εξόρυξης δεδομένων στη βιομηχανία της υγείας..58

**ΠΑΡΑΡΤΗΜΑ**

Πίνακας 1:.....78

# ΚΕΦΑΛΑΙΟ 1: Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

## 1.1 Εισαγωγή

Η εξόρυξη δεδομένων αποτελεί ένα σημαντικό εργαλείο εξεύρεσης πληροφοριών ή προτύπων από μεγάλες βάσεις δεδομένων συνδυάζοντας αρχές της στατιστικής, της τεχνητής νοημοσύνης, της μηχανής εκμάθησης και των συστημάτων βάσεων δεδομένων. Η επιστήμη της εξόρυξη δεδομένων έχει ευρύ φάσμα εφαρμογών. Στο παρόν κεφάλαιο αναλύεται η έννοια της εξόρυξης δεδομένων, παρατίθενται τα βασικά βήματα της διαδικασίας και αναλύονται οι βασικοί αλγόριθμοι που χρησιμοποιούνται. Επίσης, παρουσιάζονται τα μοντέλα πληροφοριών που εξάγονται από τον εξόρυξη δεδομένων και αναφέρονται οι βασικοί στόχοι της διαδικασίας. Τέλος, γίνεται ο βασικός διαχωρισμός στην κατευθυνόμενη και τη μη-κατευθυνόμενη ή ελεύθερη εξόρυξη δεδομένων.

## 1.2. Έννοιες – Ορισμοί

Τα τελευταία χρόνια, οι δυνατότητές μας για τη δημιουργία και τη συλλογή δεδομένων έχουν αυξηθεί ταχύτατα. Παράγοντες που συνετέλεσαν στην ανάπτυξη αυτή αποτελούν η διαδεδομένη χρήση των ραβδωτών κωδικών (bar codes) για τα περισσότερα εμπορικά προϊόντα, η μηχανογράφηση πολλών επιχειρήσεων και των δημόσιων συναλλαγών και η εξέλιξη των εργαλείων συλλογής δεδομένων. Πρέπει, επίσης, να σημειωθεί ότι ο αριθμός αυτών των βάσεων δεδομένων αυξάνεται συνεχώς, λόγω της διαθεσιμότητας των συνεχώς εξελισσόμενων και οικονομικά προσιτών συστημάτων (Chen et al. 1996). Αυτή η εκρηκτική ανάπτυξη των βάσεων δεδομένων έχει δημιουργήσει μια επείγουσα ανάγκη για τη δημιουργία νέων τεχνικών και εργαλείων, τα οποία μπορούν έξυπνα και αυτόματα να μετατρέψουν τα πρωτογενή δεδομένα σε χρήσιμες πληροφορίες και γνώσεις. Ως εκ τούτου, η εξόρυξη δεδομένων έχει καταστεί ένα ερευνητικό πεδίο με ολοένα και μεγαλύτερη σημασία.

Η εξόρυξη δεδομένων είναι η διαδικασία εξαγωγής έμμεσων, παλαιότερα γνωστών και δυνητικά χρήσιμων πληροφοριών από τα δεδομένα της βάσης δεδομένων (Frawley et al. 1992). Σύμφωνα με τους Berry και Linoff (1999), η εξόρυξη δεδομένων είναι η διαδικασία της εξερεύνησης και ανάλυσης πολλών δεδομένων, με αυτόματα ή

ημιαυτόματα μέσα, προκειμένου να ανακαλυφθούν σημαντικά πρότυπα και κανόνες. Αυτές οι νέες πληροφορίες που προκύπτουν μπορούν να χρησιμοποιηθούν σε διάφορους τομείς, όπως για παράδειγμα στην υποστήριξη της λήψης αποφάσεων, στις προβλέψεις και στις εκτιμήσεις σημαντικών επιχειρηματικών αποφάσεων, τομείς οι οποίοι μπορούν να βοηθήσουν μία επιχείρηση να αποκτήσει και να διατηρήσει ανταγωνιστικό πλεονέκτημα (Ahmed 2004).

Υπάρχουν, επίσης, πολλοί άλλοι όροι που εμφανίζονται σε ορισμένα άρθρα (Fayyad et al. 1996, Frawley et al. 1992, Shaw et al. 2001) οι οποίοι έχουν την ίδια ή ελαφρώς διαφορετική έννοια. Τέτοιοι όροι είναι οι ακόλουθοι: ανακάλυψη γνώσης σε βάσεις δεδομένων (knowledge discovery in databases), εξόρυξη γνώσης από βάσεις δεδομένων (knowledge mining from databases), εξαγωγή γνώσης (knowledge extraction), αρχαιολογία δεδομένων (data archaeology), εκβάθυνση δεδομένων (data dredging), ανάλυση δεδομένων (data analysis) και άλλα.

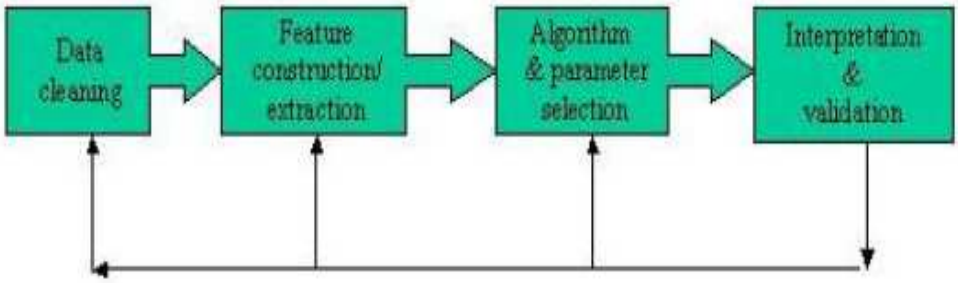
Ανέκαθεν, διάφοροι αναλυτές ασχολούνταν με την διαδικασία της εξαγωγής χρήσιμων πληροφοριών από καταγεγραμμένα δεδομένα. Στις σύγχρονες, όμως, επιχειρήσεις, ο αυξημένος όγκος των δεδομένων απαιτεί η εξαγωγή γνώσης να γίνεται με χρήση υπολογιστικών προσεγγίσεων. Πράγματι, στην σύγχρονη εποχή, ο όγκος των δεδομένων αυτών, τα οποία ονομάζονται σύνολα δεδομένων (datasets), έχει αυξηθεί τόσο σε μέγεθος, όσο και σε πολυπλοκότητα. Το αποτέλεσμα είναι οι παραδοσιακές μέθοδοι απ’ ευθείας ανάλυσης να δίνουν την θέση τους σε αυτόματες μεθόδους ανάλυσης, οι οποίες χρησιμοποιούν πιο πολύπλοκα και εξεζητημένα εργαλεία. Η μέθοδος της εξόρυξης δεδομένων χρησιμοποιεί τα δεδομένα των επιχειρήσεων ως "πρώτη ύλη" και χρησιμοποιώντας έναν προκαθορισμένο αλγόριθμο ομαδοποιεί τις τεράστιες ποσότητες αυτών σύμφωνα με τα επιθυμητά κριτήρια που μπορεί να είναι χρήσιμα για το μελλοντικό στοχευμένο μάρκετινγκ (Ahmed 2004). Η εξόρυξη δεδομένων έχει πρωτεύοντα ρόλο σε κάθε όψη της διαχείρισης πελατειακών σχέσεων. Μέσα από τις εφαρμογές των τεχνικών εξόρυξης γνώσης μπορεί μία μεγάλη επιχείρηση να μετατρέψει τις χιλιάδες εγγραφές στις βάσεις δεδομένων των πελατών της σε κάποια συνεκτικού είδους εικόνα για τους πελάτες της (Chen et al. 1996).

## 1.3 Τα βήματα της εξόρυξης δεδομένων

Σύμφωνα με τους Wu, Yu, Piatetsky-Shapiro, Cercone, Lin, Kotagiri, και Wah (2003), η διαδικασία της εξόρυξης δεδομένων αποτελείται από μία σειρά βημάτων, τα οποία είναι τα εξής:

1. Καθαρισμός των δεδομένων (data cleaning): Είναι η διαδικασία ανίχνευσης και διόρθωσης (ή αφαίρεσης) ανακριβών, λανθασμένων ή ελλιπών στοιχείων από τη βάση δεδομένων.
2. Δημιουργία/εξαγωγή των χαρακτηριστικών (feature construction/extraction): Αφορά στην απλούστευση των πόρων που απαιτούνται για να περιγραφεί ένα σύνολο δεδομένων με ακρίβεια.
3. Επιλογή αλγορίθμου και παραμέτρων (algorithm and parameter selection): Γίνεται η επιλογή του κατάλληλου αλγορίθμου εξόρυξης δεδομένων και ο καθορισμός των απαιτούμενων παραμέτρων.
4. Η ερμηνεία και επικύρωση (interpretation and validation): Αφορά στην ερμηνεία των αποτελεσμάτων και η επαλήθευση ότι τα δεδομένα πληρούν τους περιορισμούς που τέθηκαν.

Επί του παρόντος, το μεγαλύτερο ερευνητικό έργο εστιάζεται στην ανάπτυξη νέων αλγορίθμων και στη βελτίωση της ταχύτητας και της ακρίβειας των υφιστάμενων αλγορίθμων (Wu et al. 2003). Τα παραπάνω βήματα απεικονίζονται στην *εικόνα 1.1*.




Εικόνα 1.1: Η διαδικασία της εξόρυξης δεδομένων (Wu et al. 2003).

## 1.4 Οι βασικοί αλγόριθμοι της εξόρυξης δεδομένων

Οι κύριοι τύποι αλγορίθμων που χρησιμοποιούνται στην εξόρυξη δεδομένων είναι οι εξής (Ahmed 2004):

- Τα τεχνητά νευρωνικά δίκτυα (artificial neural networks): Είναι μη γραμμικά προγνωστικά μοντέλα, τα οποία μαθαίνουν μέσω εκπαίδευσης και δομικά μοιάζουν με βιολογικά νευρωνικά δίκτυα.

- Τα δέντρα αποφάσεων (decision trees): Είναι δεντρικές δομές που περιέχουν σύνολα αποφάσεων. Οι αποφάσεις αυτές παράγουν κανόνες, με σκοπό την ταξινόμηση κάποιου συνόλου δεδομένων.

- Οι γενετικοί αλγόριθμοι (genetic algorithms): Είναι αλγόριθμοι βελτιστοποίησης, οι οποίοι χρησιμοποιούν τεχνικές όπως οι γενετικοί συνδυασμοί, η μετάλλαξη και η φυσική επιλογή.

- Η επαγωγή κανόνων (rule induction): Είναι η εξαγωγή κανόνων από δεδομένα χρήσιμων if - then κανόνων, με βάση τη στατιστική τους σπουδαιότητα.

- Η μέθοδος του κοντινότερου γείτονα (nearest neighbor method): Η τεχνική αυτή ταξινομεί κάθε αντικείμενο κάποιου συνόλου δεδομένων, με βάση κάποιο συνδυασμό των κλάσεων των πιο όμοιων αντικειμένων. Συχνά ονομάζεται και k-nearest neighbor τεχνική.

- Η απεικόνιση δεδομένων (data visualization): Γίνεται οπτική ερμηνεία πολύπλοκων συναφειών σε πολυδιάστατα δεδομένα. Για τις απεικονίσεις των συναφειών χρησιμοποιούνται εργαλεία γραφικών.

## 1.5 Ο σκοπός και τα μοντέλα πληροφοριών της εξόρυξης δεδομένων

Σύμφωνα με τον Ahmed (2004), η εξόρυξη δεδομένων χρησιμοποιείται για τέσσερις κύριους σκοπούς:

1. Τη βελτίωση της διαδικασίας προσέλκυσης πελατών και τη διατήρησή τους.
2. Την ελαχιστοποίηση της απάτης.
3. Τον προσδιορισμό της εσωτερικής αναποτελεσματικότητας και στη συνέχεια, την αναδιοργάνωση των λειτουργιών της επιχείρησης.
4. Τον προσδιορισμό του ανεξερεύνητου εδάφους του διαδικτύου.

Σύμφωνα με τους Fayyad, Piatetsky-Shapiro και Smyth (1996) και τον (Ahmed 2004), η εξόρυξη δεδομένων παράγει τα ακόλουθα είδη μοντέλων πληροφοριών:

- Συσχετίσεις (associations): Οι συσχετίσεις συμβαίνουν όταν οι μεταβλητές συνδέονται μεταξύ τους με ένα μεμονωμένο γεγονός.

- Ακολουθίες (sequences): Οι ακολουθίες σχηματίζονται από δεδομένα που σχετίζονται με την έννοια του χρόνου.

- Κατάταξη (classifications): Η κατάταξη είναι η πιο διαδεδομένη δραστηριότητα εξόρυξης δεδομένων σήμερα. Συνεισφέρει στην ανακάλυψη των χαρακτηριστικών των πελατών που είναι πιθανόν να στραφούν σε ανταγωνιστές, και παρέχει ένα μοντέλο, το οποίο μπορεί να προβλέψει ποιοι είναι αυτοί οι πελάτες. Μπορεί, επίσης, να βοηθήσει στον καθορισμό των προσφορών που προσελκύουν πελάτες, αλλά και τους τύπους των πελατών που προσελκύονται, έτσι ώστε να μη δαπανώνται περισσότερα χρήματα απ’ όσα απαιτούνται για τη διατήρηση ενός πελάτη.

- Συσταδοποίηση (clustering): Χρησιμοποιώντας την συσταδοποίηση, η εξόρυξη δεδομένων ανακαλύπτει διαφορετικές ομάδες μέσα στα δεδομένα. Μπορεί να εφαρμοστεί σε αρκετά πολύπλοκα προβλήματα, όπως για παράδειγμα η ανίχνευση ελαττωματικών προϊόντων κατά την παραγωγική διαδικασία.

- Πρόγνωση (forecasting): Όλες οι παραπάνω εφαρμογές μπορεί να περιλαμβάνουν την έννοια της πρόβλεψης, όπως για παράδειγμα αν ένας πελάτης θα ανανεώσει τη συνδρομή του. Η πρόγνωση, όμως, είναι μία διαφορετική μορφή πρόβλεψης, διότι εκτιμά την μελλοντική αξία των συνεχών μεταβλητών, όπως για παράδειγμα οι πωλήσεις, με βάση μοτίβα και πρότυπα που δημιουργούνται ανάμεσα στα δεδομένα.

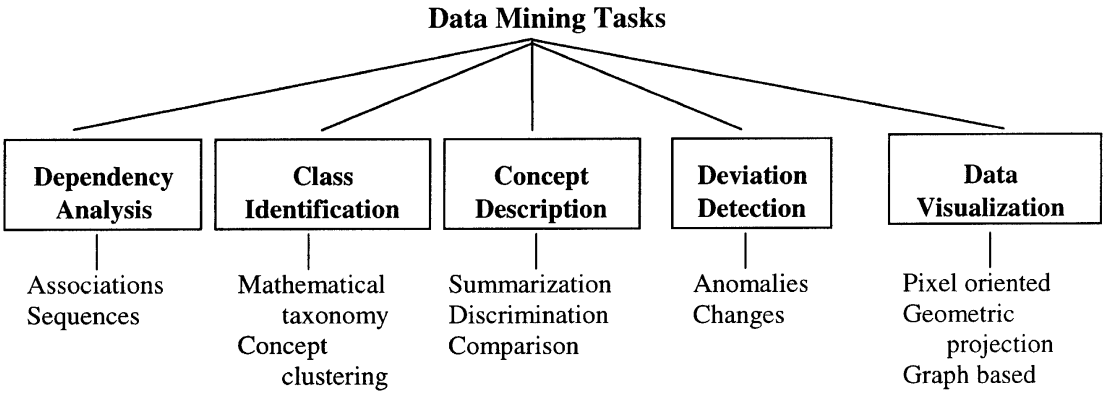
- Παλινδρόμηση (regression): Είναι η λειτουργία η οποία εισάγει ένα δεδομένο-στοιχείο σε μία πραγματικής αξίας μεταβλητή πρόβλεψης. Μία εφαρμογή της παλινδρόμησης, για παράδειγμα, είναι η ζήτηση των καταναλωτών για ένα νέο προϊόν, ως συνάρτηση της διαφημιστικής δαπάνης.

- Περιληπτική παρουσίαση (summarization): Περιλαμβάνει μεθόδους για την εύρεση μιας περιεκτικής περιγραφής για ένα υποσύνολο των δεδομένων. Ένα απλό παράδειγμα είναι η καταχώρηση σε πίνακες των μέσων τιμών και των τυπικών αποκλίσεων των δεδομένων. Οι τεχνικές περιληπτικής παρουσίασης εφαρμόζονται συχνά στην διαδραστική διερευνητική ανάλυση των δεδομένων και στην αυτοματοποιημένη δημιουργία αναφορών-εκθέσεων.

- Μοντελοποίηση εξαρτήσεων (dependency modeling): Στηρίζεται στην εύρεση κάποιου μοντέλου, το οποίο περιγράφει τις σημαντικές εξαρτήσεις μεταξύ των μεταβλητών.

- Ανίχνευση μεταβολών και αποκλίσεων (change and deviation detection): Επικεντρώνεται στην ανακάλυψη των πιο σημαντικών μεταβολών των τιμών των δεδομένων σε σχέση με προηγούμενες μετρήσεις ή κανονιστικές αξίες.

Οι Shaw, Subramaniam, Tan και Welge (2001) πραγματοποιούν μία διαφορετική διάκριση των κύριων μοντέλων πληροφοριών της εξόρυξης δεδομένων, τα οποία απεικονίζονται στην *εικόνα 1.2*. Η ταξινόμηση αυτή αντικατοπτρίζει τον αναδυόμενο ρόλο της οπτικοποίησης των δεδομένων ως ξεχωριστό έργο της εξόρυξης δεδομένων, παρότι χρησιμοποιείται για τη στήριξη άλλων λειτουργιών της εξόρυξης.




Εικόνα 1.2: Ταξινόμηση των μοντέλων πληροφοριών της εξόρυξης δεδομένων (Shaw et al. 2001).

Όπως φαίνεται και στην *εικόνα 1.2*, οι Shaw, Subramaniam, Tan και Welge (2001) ταξινομούν τα μοντέλα πληροφοριών της εξόρυξης δεδομένων σε πέντε κατηγορίες, οι οποίες είναι οι εξής:

- η ανάλυση εξαρτήσεων (dependency analysis): ο κύριος τύπος γνώσης είναι η σύνδεση των στοιχείων που έχουν κάποια κοινά χαρακτηριστικά μεταξύ τους. Μπορεί χρησιμοποιηθεί στην ανάπτυξη στρατηγικών μάρκετινγκ για την προώθηση των προϊόντων που έχουν σχέσεις εξάρτησης στην αντίληψη των πελατών.

- ο προσδιορισμός των κλάσεων (class identification): κατηγοριοποιεί τους πελάτες σε κατηγορίες, οι οποίες έχουν καθοριστεί εκ των προτέρων. Υπάρχου δύο τύποι προσδιορισμού κλάσεων, η μαθηματική ταξινόμηση (mathematical taxonomy) και η συσταδοποίηση κατά έννοια (concept clustering).

- η περιγραφή των γενικών ιδεών (concept description): είναι μία τεχνική ομαδοποίησης των πελατών με βάση την υπάρχουσα γνώση του χώρου και την βάση δεδομένων.

- η ανίχνευση αποκλίσεων (deviation detection): χρησιμεύει στην ανακάλυψη ανωμαλιών και μεταβολών στα δεδομένων.

- η οπτικοποίηση-απεικόνιση των δεδομένων (data visualization): επιτρέπει την απεικόνιση πολύπλοκων σχηματισμών στα δεδομένα των πελατών ως οπτικά αντικείμενα, σε διαστάσεις και χρώματα.

## 1.6 Ο βασικός διαχωρισμός της εξόρυξης δεδομένων

### 1.6.1 Η κατευθυνόμενη εξόρυξη δεδομένων

Σύμφωνα με τους Berry και Linoff (1999), στην κατευθυνόμενη εξόρυξη δεδομένων (directed data mining) ο στόχος μας είναι να χρησιμοποιηθούν τα διαθέσιμα δεδομένα ώστε να κατασκευαστεί ένα μοντέλο, το οποίο θα περιγράφει μία συγκεκριμένη μεταβλητή ενδιαφέροντος υπό το πρίσμα των υπόλοιπων διαθέσιμων δεδομένων. Παραδείγματα της κατευθυνόμενης εξόρυξης δεδομένων είναι τα ακόλουθα:

- Κατάταξη (classification): Η εξέταση των χαρακτηριστικών ενός νέου αντικειμένου και η τοποθέτησή του σε μία προκαθορισμένη υπάρχουσα κατηγορία.
- Εκτίμηση (estimation): Ο υπολογισμός μίας τιμής για μια άγνωστη συνεχή μεταβλητή, με βάση τα δεδομένα εισόδου.
- Πρόβλεψη (prediction): Η πρόβλεψη, μέσω ενός μοντέλου, άγνωστων ή μη διαθέσιμων τιμών.

### 1.6.2 Η μη-κατευθυνόμενη εξόρυξη δεδομένων

Στη μη-κατευθυνόμενη ή ελεύθερη εξόρυξη δεδομένων (undirected data mining), δεν επιλέγεται κάποια μεταβλητή ως στόχος. Ο σκοπός του είναι να δημιουργηθεί κάποια σχέση ανάμεσα σε όλες τις μεταβλητές. Παραδείγματα της μη-κατευθυνόμενης εξόρυξης δεδομένων είναι τα ακόλουθα (Berry and Linoff 1999):

- Ομαδοποίηση ή κανόνες συσχέτισης (affinity grouping or association rules): Ο καθορισμός των αντικειμένων που συμβαδίζουν μεταξύ τους.
- Συσταδοποίηση (clustering): Η κατάτμηση μίας ποικιλόμορφης ομάδας αντικειμένων σε συναφείς υποομάδες ή συστάδες (clusters).

- Περιγραφή και απεικόνιση (description and visualization): Η αναλυτική περιγραφή των αντικειμένων μίας βάσης δεδομένων με σκοπό τη βαθύτερη κατανόηση των δεδομένων και την απεικόνισή τους με τρόπο κατανοητό από τον ανθρώπινο παράγοντα.

# ΚΕΦΑΛΑΙΟ 2: Η ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ (CRM)

## 2.1 Εισαγωγή

Η ανάγκη για ποιοτικότερη διαχείριση των σχέσεων εταιρίας – πελάτη οδήγησε στην ανάπτυξη των συστημάτων διαχείρισης πελατειακών σχέσεων. Στο κεφάλαιο αυτό δίνεται ο ορισμός και η ερμηνεία της διαχείρισης πελατειακών σχέσεων και αναλύονται τα βήματα για την ανάπτυξη ενός αποτελεσματικού CRM συστήματος. Στη συνέχεια, περιγράφονται οι τέσσερις διαστάσεις του CRM και αναφέρονται οι τρόποι συλλογής των απαραίτητων δεδομένων των πελατών. Η διαχείριση πελατειακών σχέσεων περιλαμβάνει την τμηματοποίηση πελατών και τη δημιουργία προφίλ πελατών, διαδικασίες απαραίτητες για την ύπαρξη μίας βάσης επικοινωνίας με τους πελάτες. Επίσης, αναφέρονται οι δυσκολίες που ενδέχεται να προκύψουν κατά τη διαδικασία της τμηματοποίησης και τα απαραίτητα χαρακτηριστικά προς συλλογή για τη δημιουργία προφίλ πελατών. Τέλος, γίνεται ο διαχωρισμός του CRM στο αναλυτικό CRM και το λειτουργικό CRM.

## 2.2 Έννοιες – Ορισμοί

Η διαχείριση πελατειακών σχέσεων (Customer Relationship Management - CRM) είναι η διαδικασία συλλογής δεδομένων που αφορούν τους πελάτες μίας επιχείρησης, έτσι ώστε να γίνουν κατανοητά τα χαρακτηριστικά των πελατών και να εφαρμοστούν οι κατάλληλες στρατηγικές μάρκετινγκ. Το CRM δεν αποτελεί καινούρια έννοια, υπήρχε από το παρελθόν. Ωστόσο, την τελευταία δεκαετία, μονάχα, έχει γίνει το κέντρο της προσοχής των επιχειρήσεων (Ahn et al. 2003). Σύμφωνα με τον Bose (2002), το CRM είναι η συνένωση των τεχνολογιών και των επιχειρηματικών διαδικασιών, προκειμένου να ικανοποιηθούν οι ανάγκες του πελάτη κατά τη διάρκεια οποιασδήποτε αλληλεπίδρασης μεταξύ του πελάτη και της επιχείρησης. Κατ’ επέκταση, το CRM περιλαμβάνει την απόκτηση, την ανάλυση και την αξιοποίηση της γνώσης που

προέρχεται από τους πελάτες, προκειμένου να γίνουν περισσότερο αποτελεσματικές οι πωλήσεις των παρεχόμενων προϊόντων ή υπηρεσιών.

Ο ορισμός που δίνουν οι Parvatiyar και Sheth στο άρθρο τους (2001), είναι ο εξής: «Η διαχείριση πελατειακών σχέσεων είναι η ολοκληρωμένη στρατηγική και διαδικασία απόκτησης, διατήρησης και συνεργασίας με επιλεγμένους πελάτες με σκοπό τη δημιουργία ανώτερης αξίας για την εταιρεία και τον πελάτη. Περιλαμβάνει την ενσωμάτωση των πωλήσεων, του μάρκετινγκ, της εξυπηρέτησης πελατών και τις λειτουργίες της αλυσίδας εφοδιασμού του οργανισμού, έτσι ώστε να επιτυγχάνεται η μεγαλύτερη δυνατή αποδοτικότητα και αποτελεσματικότητα στην παροχή αξίας στους πελάτες».

Η σημερινή αγορά χαρακτηρίζεται ως παγκοσμιοποιημένη, τα προϊόντα και οι υπηρεσίες είναι σχεδόν πανομοιότυπα, ενώ υπάρχει πληθώρα προμηθευτών. Λόγω του μεγέθους και της πολυπλοκότητας των αγορών, η μαζική στόχευση των καταναλωτών είναι ιδιαίτερα ακριβή και η αποδοτικότητα των επενδυομένων κεφαλαίων είναι συχνά υπό αμφισβήτηση (Bounsaythip and Runsala 2001). Έτσι, οι επιχειρήσεις στρέφονται σε πελατοκεντρικές στρατηγικές. Ειδικότερα, η αναγνώριση των αναγκών των πελατών και η συνεχής παροχή προστιθέμενων υπηρεσιών είναι κρίσιμοι παράγοντες για την επιτυχία ή την αποτυχία των εταιρειών (Ahn et al. 2003).

Μία πελατοκεντρική εταιρεία είναι σε θέση να αντιμετωπίσει κάθε πελάτη ξεχωριστά και μοναδικά, ανάλογα με την προτίμηση του πελάτη. Σύμφωνα με τους Berger και Bechwati (2000), το επίκεντρο της διαχείρισης πελατειακών σχέσεων είναι η ανάπτυξη και διατήρηση μακροχρόνιων σχέσεων με τους πελάτες, και όχι απλώς μια σειρά από ξεχωριστές συναλλαγές. Υποστηρίζουν, επίσης, ότι η επιχείρηση θα πρέπει να εξετάσει την αξία του πελάτη σύμφωνα με τη συνολική του σχέση με την εταιρεία, και όχι να υπολογίζει το κέρδος από κάθε ξεχωριστή συναλλαγή. Οι πιο σημαντικές λειτουργίες μιας πελατοκεντρικής στρατηγικής μάρκετινγκ είναι η συλλογή και η συσσώρευση πληροφοριών για τους πελάτες και η παροχή υπηρεσιών σε αυτούς. Η συλλογή των απαιτούμενων πληροφοριών καθίσταται δύσκολη, εάν η εταιρεία δεν εκμεταλλευθεί κατάλληλα την τεχνολογία των πληροφοριών (Information Technology). Οι ραγδαίες εξελίξεις της τεχνολογίας, όπως οι βάσεις δεδομένων, η αποθήκευση των δεδομένων (data warehousing) και η εξόρυξη δεδομένων παίζουν κυρίαρχο ρόλο στις νέες στρατηγικές μάρκετινγκ (Ahn et al. 2003).

## 2.3 Τα στάδια ανάπτυξης ενός CRM συστήματος

Πέρα από τη σύνθετη ολοκλήρωση του υλικού (hardware), του λογισμικού (software) και των εφαρμογών (applications), η δημιουργία ενός CRM συστήματος απαιτεί την διεξοδική ανάλυση των επιχειρηματικών διαδικασιών. Η απαιτούμενη εργασία, για να προσαρμοστεί ένα τέτοιο σύστημα στην πραγματικότητα, απαιτεί μεγάλη ποικιλία γνώσης, σωστή διαχείριση του έργου και ένα εμπεριστατωμένο σχέδιο. Σύμφωνα με τον Bose (2002), η ανάπτυξη ενός CRM συστήματος αποτελείται από οκτώ βήματα-στάδια, τα οποία αναλύονται παρακάτω.

Στάδιο 1. Σχεδιασμός (planning): Πρωτίστως, η εταιρεία θα πρέπει να προσδιορίσει το πώς, πότε και πού θα έρχεται σε επαφή με τους πελάτες της. Αφού έχει προσδιορίσει όλα τα σημεία αλληλεπίδρασης, θα πρέπει στη συνέχεια να αποφασίσει εάν θα διατηρήσει, θα τροποποιήσει ή θα καταργήσει αυτά τα σημεία επαφής. Ο υπεύθυνος των πληροφοριακών συστημάτων (IT manager) θα πρέπει να εστιάσει στο πώς αυτές οι αλληλεπιδράσεις μπορούν να καταγραφούν σε ένα πληροφοριακό σύστημα. Ο δεύτερος παράγοντας μελέτης της επιχείρησης αφορά στην υποστήριξη της λήψης αποφάσεων, δηλαδή το πώς οι παρεχόμενες πληροφορίες μπορούν να βελτιώσουν την ποιότητα των αποφάσεων, έτσι ώστε να συμβαδίζουν με την πελατοκεντρική προσέγγιση που έχει υιοθετηθεί.

Στάδιο 2. Έρευνα (research): Στο στάδιο αυτό, οι άνθρωποι των πληροφοριακών συστημάτων θα πρέπει να εντοπίσουν τρόπους για την αντιμετώπιση των αναγκών του οργανισμού στο πλαίσιο του CRM συστήματος. Θα πρέπει, βεβαίως, να ληφθούν υπόψιν η οργανωσιακή δομή της εταιρείας, η κουλτούρα της, οι δυνατοί προμηθευτές λογισμικού, και άλλοι παράγοντες που συμβάλλουν στην ιδιαιτερότητα της κάθε επιχείρησης.

Στάδιο 3. Ανάλυση συστήματος (system analysis): Αποτελεί το πιο σημαντικό στάδιο, σε συνδυασμό με το στάδιο της έρευνας. Μερικοί κρίσιμοι παράγοντες που θα πρέπει να ληφθούν υπόψιν σε αυτό το στάδιο είναι οι εξής: η απόκτηση εξωτερικών εμπειρογνωμόνων, η εξέταση ενός στρατηγικού σχεδίου, ο εκ νέου σχεδιασμός των δεδομένων των πελατών, η προσαρμογή των δεδομένων ώστε να είναι κατάλληλα για τη λήψη αποφάσεων, η δυνατότητα επεκτασιμότητας, και τέλος, η μελέτη σκοπιμότητας.

Στάδιο 4. Σχεδιασμός (design): Αφορά τον λεπτομερή καθορισμό των πακέτων λογισμικού, σε συνδυασμό με τις βασικές τεχνολογικές υποδομές, όπως οι αποθήκες δεδομένων, η αρχιτεκτονική του δικτύου και άλλα. Στο στάδιο αυτό, η γνώμη ενός έμπειρου συμβούλου μπορεί να είναι απαραίτητη, ώστε να καθοριστούν οι αλλαγές και οι τροποποιήσεις, εφόσον αυτές κρίνονται απαραίτητες, προκειμένου να συνδεθούν αρμονικά όλα τα συστήματα.

Στάδιο 5. Κατασκευή (construction): Είναι η εκτέλεση του σταδίου του σχεδιασμού και αποτελεί ένα αρκετά χρονοβόρο έργο. Για το λόγο αυτό, εκτός και αν το έργο έχει διαιρεθεί σε επιμέρους στάδια, πολλές εταιρείες βασίζονται σε έναν συγκεκριμένο προμηθευτή ή σύμβουλο, ώστε να εξασφαλιστεί η ομαλή μετάβαση.

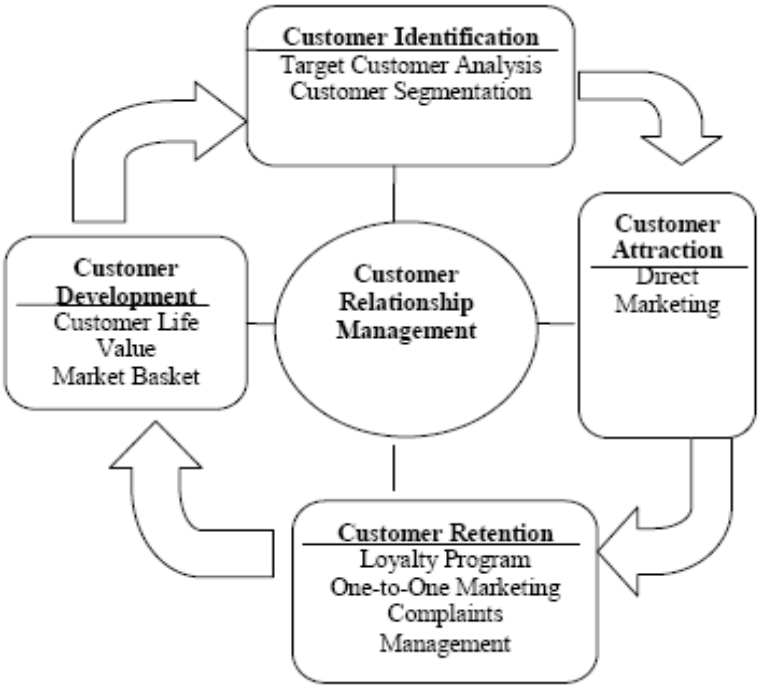
Στάδιο 6. Εκτέλεση (implementation): Ένα κρίσιμο στοιχείο του σταδίου αυτού είναι η κατάλληλη εκπαίδευση του προσωπικού, έτσι ώστε να είναι σε θέση να κατανοήσει πλήρως και να χειριστεί το νέο σύστημα. Ενώ μερικά από τα εργαλεία που χρησιμοποιούνται στο CRM σύστημα φαίνεται ότι παράγουν πλήρως αξιοποιήσιμες πληροφορίες, τα αποτελέσματα δεν πρέπει να θεωρούνται δεδομένα.

Στάδιο 7. Συντήρηση και τεκμηρίωση (maintenance and documentation): Αποτελεί ένα εξίσου σημαντικό στάδιο, αφού η επιχείρηση θα πρέπει πάντα να επιδιώκει να αποκτά περισσότερες πληροφορίες για τους πελάτες της. Λόγω της δυναμικότητας της αγοράς, το CRM σύστημα απαιτεί τη συνεχή αξιολόγηση της απόδοσης του συστήματος. Οι άνθρωποι των πληροφοριακών συστημάτων θα πρέπει να συνεργάζονται και με άλλους λειτουργικούς τομείς, όπως το μάρκετινγκ, η διοίκηση και η παραγωγή, ώστε να διασφαλιστεί ότι το σύστημα ικανοποιεί τις ανάγκες των φορέων λήψης αποφάσεων στην επιχείρηση.

Στάδιο 8. Προσαρμογή (adaptation): Το στάδιο αυτό θεωρείται απαραίτητο, ενόσω το CRM σύστημα βρίσκεται ακόμα σε αρχικό στάδιο. Εάν δεν γίνονται οι απαραίτητες προσαρμογές σύμφωνα με τις νεοεισερχόμενες πληροφορίες, η επιχείρηση θα χάσει γρήγορα το ανταγωνιστικό πλεονέκτημα της πελατοκεντρικής της προσέγγισης.

## 2.4 Οι τέσσερις διαστάσεις του CRM

Τα προγράμματα CRM, όταν σχεδιάζονται προσεκτικά και εφαρμόζονται με επιτυχία, υπόσχονται πολλά προνόμια, τόσο για τους πελάτες όσο και για τις επιχειρήσεις. Το CRM αποτελείται από τέσσερις διαστάσεις: (1) Εντοπισμός πελατών (customer identification), (2) Προσέλκυση πελατών (customer attraction), (3) Διατήρηση πελατών (customer retention), και (4) Ανάπτυξη πελατών (customer development) (Liu and Zhu 2009). Οι τέσσερις αυτές διαστάσεις απεικονίζονται στην *εικόνα 2.1*.




Εικόνα 2.1 : Οι τέσσερις διαστάσεις του CRM (Liu and Zhu 2009).

Οι τέσσερις διαστάσεις του CRM αναλύονται ως εξής:

1. Εντοπισμός πελατών (customer identification). Περιλαμβάνει τον εντοπισμό των ατόμων που έχουν τις περισσότερες πιθανότητες να γίνουν πελάτες ή θα είναι οι περισσότερο επικερδείς για την εταιρεία. Επιπλέον, περιλαμβάνει την ανάλυση των πελατών που έχουν στραφεί σε ανταγωνιστές, όπως επίσης και το πώς μπορεί η εταιρεία να τους ανακτήσει. Σε αυτό τα στάδιο γίνεται και η τμηματοποίηση πελατών, κατά την οποία υποδιαιρείται η πελατειακή βάση σε μικρότερες ομάδες πελατών.

2. Προσέλκυση πελατών (customer attraction). Μετά τον εντοπισμό των τμημάτων των δυνητικών πελατών, η εταιρεία μπορεί πλέον να στραφεί στον εντοπισμό των κατάλληλων πόρων και στην προσπάθεια προσέλκυσης των πελατών-στόχων. Στοιχείο της προσέλκυσης των πελατών αποτελεί το άμεσο μάρκετινγκ, δηλαδή η διαδικασία προώθησης του προϊόντος/υπηρεσίας μέσω διαφόρων καναλιών.

3. Διατήρηση των πελατών (customer retention). Αποτελεί το κεντρικό σημείο ενδιαφέροντος του CRM. Η ικανοποίηση του πελάτη αποτελεί απαραίτητη προϋπόθεση για τη διατήρησή του. Εργαλεία της διατήρησης πελατών αποτελούν το εξατομικευμένο μάρκετινγκ (one-to-one marketing), τα προγράμματα πιστότητας (loyalty programs) και η διαχείριση των παραπόνων των πελατών. Το εξατομικευμένο μάρκετινγκ αναφέρεται στις καμπάνιες μάρκετινγκ που υποστηρίζονται από την ανάλυση, τον εντοπισμό και την πρόβλεψη των αλλαγών στη συμπεριφορά των πελατών. Τα προγράμματα πιστότητας περιλαμβάνουν τις διαφημιστικές καμπάνιες και τις δραστηριότητες που στοχεύουν στη διατήρηση μιας μακροπρόθεσμης σχέση με τους πελάτες.

4. Ανάπτυξη πελατών (customer development). Στοιχεία της ανάπτυξης πελατών αποτελούν η ανάλυση διαχρονικής αξίας του πελάτη (lifetime value analysis), οι σταυροειδείς πωλήσεις (up/cross selling) και η ανάλυση καλαθιού μάρκετινγκ (market basket analysis). Η ανάλυση διαχρονικής αξίας του πελάτη ορίζεται ως η πρόβλεψη των συνολικών καθαρών εσόδων που αναμένει η εταιρεία από έναν πελάτη. Οι σταυροειδείς πωλήσεις αναφέρονται στις προωθητικές ενέργειες που στοχεύουν στη διεύρυνση των στενά συνδεδεμένων υπηρεσιών/προϊόντων που χρησιμοποιεί ο πελάτης. Τέλος, η ανάλυση καλαθιού μάρκετινγκ στοχεύει στη μεγιστοποίηση του αριθμού των συναλλαγών των πελατών, και κατ' επέκταση της αξίας τους, μέσω της τακτικής καταναλωτικής συμπεριφοράς τους.

## 2.5 Συλλογή των δεδομένων των πελατών

Η συλλογή των απαιτούμενων δεδομένων των πελατών θεωρείται προϋπόθεση για την ανάπτυξη ενός αποτελεσματικού CRM συστήματος. Υπάρχουν πολλοί τρόποι για να συλλεχθούν τα απαραίτητα δεδομένα, και αυτοί είναι οι εξής (Bounsaythip and Runsala 2001):

- Εσωτερική βάση δεδομένων πελατών. Τα ονόματα μπορούν να προέλθουν από τα άμεσα προγράμματα ηλεκτρονικής αλληλογραφίας που έχουν χρησιμοποιηθεί στο παρελθόν, τα προγράμματα αγοραστών, διαγωνισμούς, έγγραφες εγγυήσεις, αποδείξεις και κάρτες μελών.

- Εξωτερικές πηγές. Υπάρχουν λογισμικά ή βάσεις δεδομένων που μπορούν να ανακαλύψουν στοιχεία για τον τρόπο ζωής, δημογραφικές πληροφορίες, χρησιμοποιώντας για παράδειγμα μόνο τον ταχυδρομικό κώδικα, και άλλες χρήσιμες πληροφορίες.

- Ερευνητικές μελέτες. Αυτές μπορούν να διεξαχθούν πρόσωπο με πρόσωπο, τηλεφωνικώς, με ταχυδρομικά ερωτηματολόγια ή μέσω του διαδικτύου.

Τα συλλεγμένα δεδομένα, ανάλογα με τις πληροφορίες που μας διαθέτουν, χωρίζονται σε δύο κατηγορίες, τις μεταβλητές ταξινόμησης (classification variables) και τις περιγραφικές μεταβλητές (descriptor variables) (Kimbal 1997). Οι μεταβλητές ταξινόμησης χρησιμοποιούνται για την ταξινόμηση πελατών σε τμήματα. Αυτές οι μεταβλητές μπορεί να είναι δημογραφικές, ψυχογραφικές ή συμπεριφορικές μεταβλητές (Kimbal 1997, Goyat 2011).

- Δημογραφικές μεταβλητές (demographic variables): Ηλικία, φύλο, εισόδημα, οικογενειακή κατάσταση, εκπαιδευτικό επίπεδο, μέγεθος της οικογένειας, τύπος κατοικίας, κτλ..

- Γεωγραφικές μεταβλητές (geographic variables): Πόλη, ταχυδρομικός κώδικας, περιοχή, χώρα, αστική ή αγροτική τοποθεσία, πυκνότητα πληθυσμού, κλίμα, κτλ..

- Ψυχογραφικές μεταβλητές (psychographic variables): Συμπεριφορές, τρόπος ζωής, χόμπι, αποστροφή ή όχι απέναντι στον κίνδυνο, χαρακτηριστικά της προσωπικότητας, ηγετικά γνωρίσματα, περιοδικά που διαβάζονται, τηλεοπτικά προγράμματα που παρακολουθούνται, κτλ..

- Συμπεριφορικά χαρακτηριστικά (behavioral variables): Εμπιστοσύνη στη μάρκα του προϊόντος, συχνότητα χρήσης, οφέλη που επιδιώκονται, χρησιμοποιούμενα κανάλια διανομής, αντίδραση στους παράγοντες του μάρκετινγκ κτλ..

Οι περιγραφικές μεταβλητές χρησιμεύουν στην περιγραφή του κάθε τμήματος και στην διάκριση της μίας ομάδας από την άλλη. Πρέπει να είναι εύκολα μετρήσιμες, ή να συνδέονται με κάποια μέτρα που υπάρχουν ή να μπορούν να προσαρτηθούν στους φακέλους των πελατών. Πολλές από τις μεταβλητές ταξινόμησης μπορούν, επίσης, να θεωρηθούν ως περιγραφικές μεταβλητές. Ωστόσο, μόνο ένα πολύ μικρό ποσοστό από αυτές είναι άμεσα διαθέσιμες από εξωτερικές πηγές.

## 2.6 Τμηματοποίηση πελατών και δημιουργία προφίλ πελατών

Αντί να θέτονται ως στόχοι όλοι οι καταναλωτές ανεξαιρέτως και να τους παρέχεται το ίδιο κίνητρο, η εταιρεία μπορεί να επιλέξει εκείνους τους καταναλωτές που πληρούν ορισμένα κριτήρια, με βάση τις ατομικές τους ανάγκες και τις αγοραστικές τους συνήθειες. Αυτό επιτυγχάνεται με την οικοδόμηση ενός μοντέλου, το οποίο μπορεί να προβλέψει την μελλοντική αξία ενός ατόμου με βάση τα δημογραφικά του χαρακτηριστικά, τον τρόπο ζωής του, αλλά και την προγενέστερη καταναλωτική συμπεριφορά του. Αυτό το μοντέλο παράγει πληροφορίες οι οποίες εστιάζουν στην διατήρηση των πελατών αλλά και στην προσέλκυση νέων, δημιουργώντας και διατηρώντας την πλέον κερδοφόρα πελατειακή βάση. Αυτή η διαδικασία ονομάζεται «Μοντελοποίηση της Καταναλωτικής Συμπεριφοράς» (Customer Behavior Modeling - CBM) ή αλλιώς «Δημιουργία Προφίλ Πελατών» (Customer Profiling). Η δημιουργία προφίλ πελατών είναι ένα εργαλείο το οποίο συνεισφέρει στη βαθύτερη κατανόηση των χαρακτηριστικών της πελατειακής βάσης (Thearling 2000).

Η διαχείριση πελατειακών σχέσεων περιλαμβάνει την τμηματοποίηση πελατών (customer segmentation) και τη δημιουργία προφίλ πελατών (customer profiling) (Bounsaythip and Runsala 2001). Ο όρος τμηματοποίηση πελατών χρησιμοποιείται για να περιγράψει την διαδικασία διαχωρισμού των πελατών σε ομογενείς ομάδες, με βάση τα κοινά τους χαρακτηριστικά, όπως για παράδειγμα κοινές συνήθειες, προτιμήσεις κτλ.. Ο όρος δημιουργία προφίλ πελατών περιγράφει τους πελάτες σύμφωνα με τα ιδιαίτερα χαρακτηριστικά τους, όπως για παράδειγμα την ηλικία, το εισόδημα και τον τρόπο ζωής. Αυτό πραγματοποιείται δημιουργώντας ένα μοντέλο συμπεριφοράς του πελάτη και εκτιμώντας τις παραμέτρους του. Η δημιουργία προφίλ πελατών εκτελείται μετά την τμηματοποίηση πελατών.

### 2.6.1 Τμηματοποίηση των πελατών

Σύμφωνα με τον Goyat (2011), η τμηματοποίηση πελατών είναι ένας τρόπος για να υπάρξει περισσότερο στοχευμένη επικοινωνία με τους πελάτες. Η διαδικασία της τμηματοποίησης περιγράφει τα χαρακτηριστικά των πελατών και τους κατατάσσει σε ομάδες, οι οποίες ονομάζονται τμήματα (segments) ή συμπλέγματα (clusters). Η κατάτμηση (segmenting) είναι η τοποθέτηση του πληθυσμού σε τμήματα σύμφωνα με συγγενικά ή παρόμοια χαρακτηριστικά. Η τμηματοποίηση πελατών είναι ένα στάδιο προετοιμασίας για την ταξινόμηση του κάθε πελάτη στις ομάδες πελατών που έχουν οριστεί, και κρίνεται απαραίτητη προκειμένου να αντιμετωπιστούν οι σημερινές συνεχώς μεταβαλλόμενες και κατακερματισμένες καταναλωτικές αγορές. Μέσω της διαδικασίας αυτής, καθίσταται αποτελεσματικότερη η διοχέτευση των διαθέσιμων πόρων και ανακαλύπτονται περισσότερες ευκαιρίες ανάπτυξης.

Κατά τη διαδικασία της τμηματοποίησης πελατών μπορεί να προκύψουν ορισμένες δυσκολίες, οι οποίες είναι οι ακόλουθες (Bounsaythip and Runsala 2001):

- Η συνάφεια και ποιότητα των δεδομένων (relevance and quality of the data): Αυτά τα δύο στοιχεία είναι απαραίτητα για τη δημιουργία αξιοποιήσιμων τμημάτων των πελατών. Η ανεπάρκεια των δεδομένων ή ο υπερβολικά μεγάλος όγκος δεδομένων μπορεί να οδηγήσει σε πολύπλοκες ή χρονοβόρες αναλύσεις. Εάν τα δεδομένα δεν είναι καλά οργανωμένα (διαφορετικές μορφές, διαφορετικές πηγές δεδομένων), τότε όχι μόνο θα είναι δύσκολο να εξαχθούν οι κατάλληλες πληροφορίες, αλλά και η επιχείρηση δε θα μπορεί να αξιοποιήσει κατάλληλα την τμηματοποίηση πελατών που θα προκύψει. Συγκεκριμένα, η χρήση πολλών μεταβλητών μπορεί να προκαλέσει σύγχυση, και η τμηματοποίηση που θα προκύψει θα είναι ακατάλληλη για τη λήψη αποφάσεων από τα στελέχη της επιχείρησης. Αντίθετα, φαινομενικά αποτελεσματικές μεταβλητές μπορεί να μην είναι εύκολα αναγνωρίσιμες. Πολλά από αυτά τα προβλήματα οφείλονται σε ανεπαρκείς βάσεις δεδομένων των πελατών.

- Διαίσθηση (intuition): Παρόλο που τα υπάρχοντα δεδομένα μπορούν να παρέχουν πολλές πληροφορίες, θα πρέπει η επιχείρηση να αναπτύσσει συνεχώς υποθέσεις και παραδοχές για την τμηματοποίηση των πελατών, προκειμένου να εντοπίζονται τα κατάλληλα δεδομένα για περαιτέρω ανάλυση.

- Συνεχής διαδικασία (continuous process): Η σωστή τμηματοποίηση απαιτεί συνεχή ανάπτυξη και ενημέρωση, καθώς καθημερινά αποκτούνται νέα δεδομένα, και μπορεί να κριθεί απαραίτητη η αναθεώρηση και η ανακατάταξη των πελατών στα τμήματα. Συγκεκριμένα, σε ένα ηλεκτρονικά εμπορικό περιβάλλον, στο οποίο η ανατροφοδότηση γίνεται αυτόματα, η τμηματοποίηση απαιτεί μία καθημερινή ενημέρωση (update).

- Υπερ-τμηματοποίηση (over-segmentation): Ένα τμήμα μπορεί να καταλήξει πολύ μικρό ή και ανεπαρκές, ώστε να απαιτεί διαφορετική μεταχείριση, σε σχέση με τα υπόλοιπα τμήματα.

Η μέθοδος της τμηματοποίησης είναι ένας οικονομικά αποδοτικός τρόπος στόχευσης των πελατειακών ομάδων, και αποδείχτηκε στο παρελθόν ως ένα ισχυρό ανταγωνιστικό πλεονέκτημα. Ωστόσο, σήμερα, η τμηματοποίηση των πελατών δεν αποτελεί πλέον ανταγωνιστικό πλεονέκτημα, και θεωρείται ως η ελάχιστη προϋπόθεση για τη συνέχιση των επιχειρηματικών δραστηριοτήτων. Προκειμένου να ανακτηθεί το ανταγωνιστικό πλεονέκτημα, οι ηγετικές εταιρείες εισάγουν, πλέον, ένα νέο προσανατολισμό, ο οποίος θα μπορούσε να ονομαστεί πελατοκεντρικός προσανατολισμός (customer-centric orientation) (Bose 2002).

### 2.6.2 Δημιουργία προφίλ πελατών

Η δημιουργία προφίλ πελατών παρέχει μία βάση για την επικοινωνία της επιχείρησης με τους υπάρχοντες πελάτες, με στόχο τη διατήρησή τους αλλά και την προσφορά καλύτερων υπηρεσιών. Αυτό επιτυγχάνεται με την συγκέντρωση πληροφοριών για κάθε πελάτη, όπως για παράδειγμα δημογραφικά και συμπεριφορικά δεδομένα. Η διαδικασία αυτή συνεισφέρει, επίσης, στον εντοπισμό δυνητικών πελατών, χρησιμοποιώντας εξωτερικές πηγές, όπως για παράδειγμα η αγορά δημογραφικών και στατιστικών δεδομένων από εξειδικευμένες εταιρείες. Μέσω αυτών των δεδομένων καθίσταται δυνατή η κατάτμηση της πελατειακής βάσης σε ομάδες πελατών με ιδιαίτερα αγοραστικά γνωρίσματα (Wilson R. 2000).

Αναλόγως με το στόχο που έχει θέσει η κάθε εταιρεία, θα πρέπει αυτή να επιλέξει τα στοιχεία από τα οποία θα απαρτίζεται το προφίλ του πελάτη. Ένα απλοποιημένο προφίλ περιέχει τουλάχιστον τα βασικά προσωπικά του στοιχεία, δηλαδή το όνομα, τη διεύθυνση, την πόλη, το νομό και τον ταχυδρομικό κώδικα. Εάν η εταιρεία θέλει να δει πώς σχετίζεται το προφίλ του πελάτη με συγκεκριμένα προϊόντα, το αρχείο αυτό θα πρέπει να εμπεριέχει περισσότερες πληροφορίες, όπως το είδος του προϊόντος που αγόρασε, την ποσότητα του προϊόντος και το σύνολο των χρημάτων που δαπάνησε.

Τα χαρακτηριστικά των πελατών που μπορούν να χρησιμοποιηθούν στη διαδικασία δημιουργίας του προφίλ του πελάτη, είναι τα ακόλουθα (Wilson 2000):

- Γεωγραφικά: Η ομαδοποίηση γίνεται σε περιφερειακό, εθνικό ή παγκόσμιο επίπεδο;

- Πολιτιστικά και εθνικά: Τι γλώσσες μιλούν; Η εθνικότητα επηρεάζει τις προτιμήσεις ή τις καταναλωτικές συνήθειες;

- Οικονομικές συνθήκες, εισόδημα ή και αγοραστική δύναμη: Ποιο είναι το μέσο οικογενειακό εισόδημα ή η αγοραστική δύναμη του καταναλωτή; Αντιμετωπίζουν δυσκολίες πληρωμής; Πόσο συχνά και πόσα χρήματα δαπανά ο πελάτης σε κάθε προϊόν;

- Η συχνότητα αγορών, η συχνότητα εκδήλωσης παραπόνων, ο βαθμός της ικανοποίησης και οι προτιμήσεις στα προϊόντα μπορούν να χρησιμοποιηθούν στην κατασκευή ενός αγοραστικού προφίλ.

- Η ηλικία: Ποια είναι η κυρίαρχη ηλικιακή ομάδα της αγοράς-στόχου; Πόσα παιδιά βρίσκονται σε κάθε οικογένεια και τι ηλικία έχουν;

- Αξίες, συμπεριφορές, πεποιθήσεις: Ποια είναι η στάση των καταναλωτών απέναντι στο είδος του προϊόντος ή της παρεχόμενης υπηρεσίας;

- Ο κύκλος ζωής: Πόσο καιρό αγοράζει σε τακτική βάση τα προϊόντα;

- Γνώση και επίγνωση: Τι επίπεδο γνώσης έχει ο πελάτης για ένα προϊόν/υπηρεσία ή για τη βιομηχανία; Ποιο είναι το επίπεδο της εκπαίδευσης που απαιτείται; Τι ύψους

διαφήμιση απαιτείται για να έχει μία ομάδα πελατών επίγνωση της εκάστοτε προσφοράς;

- Τρόπος ζωής: Πόσα και ποια χαρακτηριστικά του τρόπου ζωής των αγοραστών μας είναι χρήσιμα;

- Χρησιμοποιούμενα μέσα ενημέρωσης: Από που πληροφορούνται οι καταναλωτές-στόχος; Τι διαβάζουν; Σε ποια περιοδικά και εφημερίδες είναι συνδρομητές;

- Μέθοδος στρατολόγησης: Πώς επέλεξε το δικό μας προϊόν/υπηρεσία ο καταναλωτής;

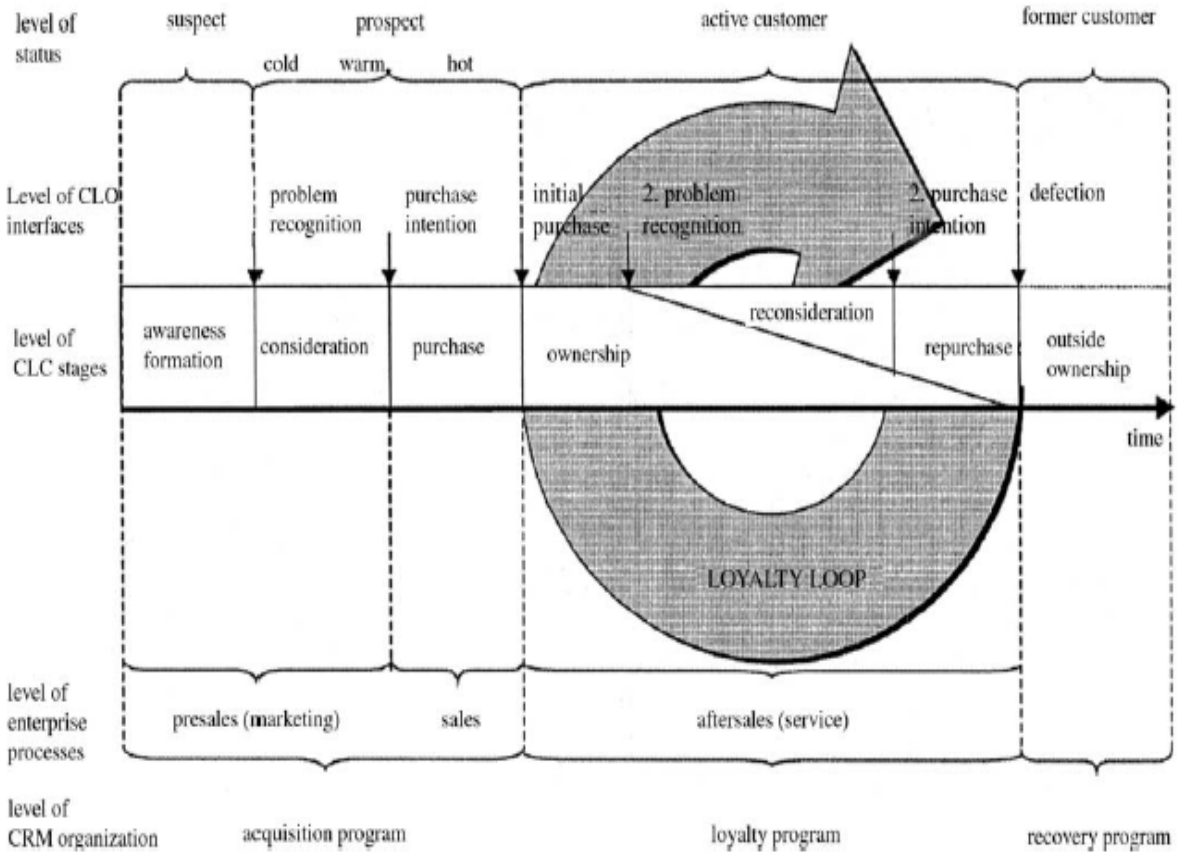
## 2.7 Επιχειρησιακό CRM και Αναλυτικό CRM

Το επιχειρησιακό CRM (operational CRM – oCRM) περιλαμβάνει τις δραστηριότητες οι οποίες σχετίζονται με την άμεση επαφή με τον πελάτη, όπως είναι οι ανοικτές γραμμές επικοινωνίας (hotlines), οι διαφημιστικές καμπάνιες ή τα κλαμπ πελατών. Κάθε oCRM δραστηριότητα συνήθως υλοποιείται σε μία από τις τρεις επιχειρησιακές διαδικασίες: τις πωλήσεις, το μάρκετινγκ ή τις παρεχόμενες υπηρεσίες, δεδομένου ότι αυτές οι διαδικασίες αφορούν την άμεση επαφή με τον πελάτη (Xu and Walton 2005). Το αναλυτικό CRM (analytical CRM – aCRM) περιλαμβάνει όλες τις συνιστώσες για την ανάλυση των χαρακτηριστικών-συμπεριφορών των πελατών, προκειμένου να επιτευχθούν οι oCRM δραστηριότητες (He et al. 2004). Επομένως, το αναλυτικό CRM είναι σε θέση να βοηθήσει την επιχείρηση να διακρίνει και να κατανέμει αποτελεσματικότερα τους διαθέσιμους πόρους στις πιο κερδοφόρες ομάδες πελατών. Η εξόρυξη δεδομένων εφαρμόζεται στο αναλυτικό CRM.

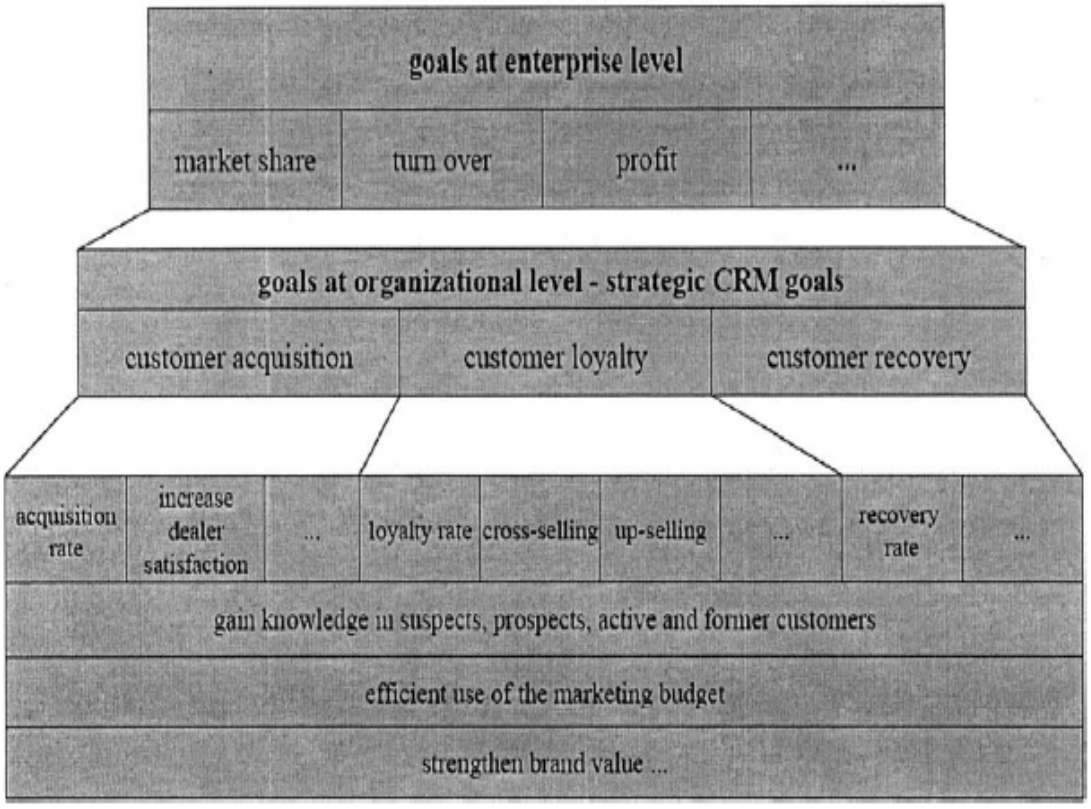
Προτού οι επιχειρήσεις μπορέσουν να αναπτύξουν στρατηγικές μάρκετινγκ ή CRM, θα πρέπει να κατανοήσουν το πώς οι καταναλωτές λαμβάνουν τις αγοραστικές τους αποφάσεις. Αυτή η διαδικασία ονομάζεται κύκλος αγοράς του πελάτη (customer buying cycle – CBC). Η αλυσίδα όλων των CBCs ενός συγκεκριμένου πελάτη διασχίζει ολόκληρο τον κύκλο ζωής του πελάτη. Η διαδικασία ολοκληρώνεται με την τελική παύση κατανάλωσης του προϊόντος–υπηρεσίας. Προκειμένου να οργανωθούν οι oCRM και οι aCRM δραστηριότητες κατά μήκος της διαδικασίας CRM, εφαρμόζονται ως

ξεχωριστά προγράμματα, με σαφείς διασυνδέσεις, ειδικούς στόχους και αντίστοιχες άμεσες ενέργειες μάρκετινγκ, όπως οι εκστρατείες απόκτησης νέων πελατών (Arndt and Gestern 2001).

Η εφαρμογή των δραστηριοτήτων oCRM και aCRM ως ξεχωριστά προγράμματα απεικονίζεται στην *εικόνα 2.2*, ενώ στη *εικόνα 2.3* απεικονίζεται η ένταξη των στόχων του CRM στην πυραμίδα γενικών στόχων της επιχείρησης.




Εικόνα 2.2: Η CRM διαδικασία με βάση τον κύκλο ζωής του πελάτη (Arndt and Gestern 2001).




Εικόνα 2.3: Ένταξη των στόχων του CRM στους στόχους της επιχείρησης (Arndt and Gestern 2001).

# ΚΕΦΑΛΑΙΟ 3: Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ

## 3.1 Εισαγωγή

Η εξόρυξη δεδομένων μπορεί να βοηθήσει τις επιχειρήσεις να ανταπεξέλθουν σε περιόδους αιχμής της κατανάλωσης, στη διακίνηση των εμπορευμάτων, αλλά και στη μείωση των ζημιών που οφείλονται σε εσωτερικές απάτες (internal fraud), οι οποίες υπολογίζεται ότι είναι περίπου το 40% έως 50 % των αποθεμάτων των επιχειρήσεων λιανικής πώλησης (Ahmed 2004). Η εξόρυξη δεδομένων μπορεί επίσης να βοηθήσει στην επίλυση ασυνήθιστων μοτίβων που αφορούν επιστροφές εμπορευμάτων, εκπτώσεις, υπερισχύουσες τιμές, πιστωτικές κάρτες, κάρτες καταστημάτων, χρεωστικές κάρτες, εκπτώσεις που παρέχονται στο προσωπικό, πλεονάζοντα εμπορεύματα και ελλείψεις που οφείλονται στα αποθέματα που αναφέρονται ως κατεστραμμένα ή

ελαττωματικά, καθιστώντας με αυτόν τον τρόπο την ανίχνευση της λιανικής απάτης πολύ πιο εύκολη, ακριβής, έγκαιρη και οικονομική.

Στο παρόν κεφάλαιο γίνεται η ταξινόμηση των τεχνικών εξόρυξης δεδομένων που εφαρμόζονται στη διαχείριση πελατειακών σχέσεων, αναφέρονται τα βήματα για την προετοιμασία των δεδομένων, προκειμένου αυτά να εισέλθουν στη διαδικασία της εξόρυξης, και αναλύονται τα στάδια για την αποτελεσματική εφαρμογή της εξόρυξης δεδομένων στο CRM. Στη συνέχεια, παρουσιάζονται το μοντέλο τμηματοποίησης πελατών και το σύστημα δημιουργίας προφίλ πελατών χρησιμοποιώντας τα εργαλεία της εξόρυξης δεδομένων. Τέλος, γίνεται αναφορά στο Marketing Data Intelligence, το οποίο είναι το αποτέλεσμα της εφαρμογής της εξόρυξης δεδομένων στον κύκλο ζωής του πελάτη.

## 3.2 Έννοιες – Ορισμοί

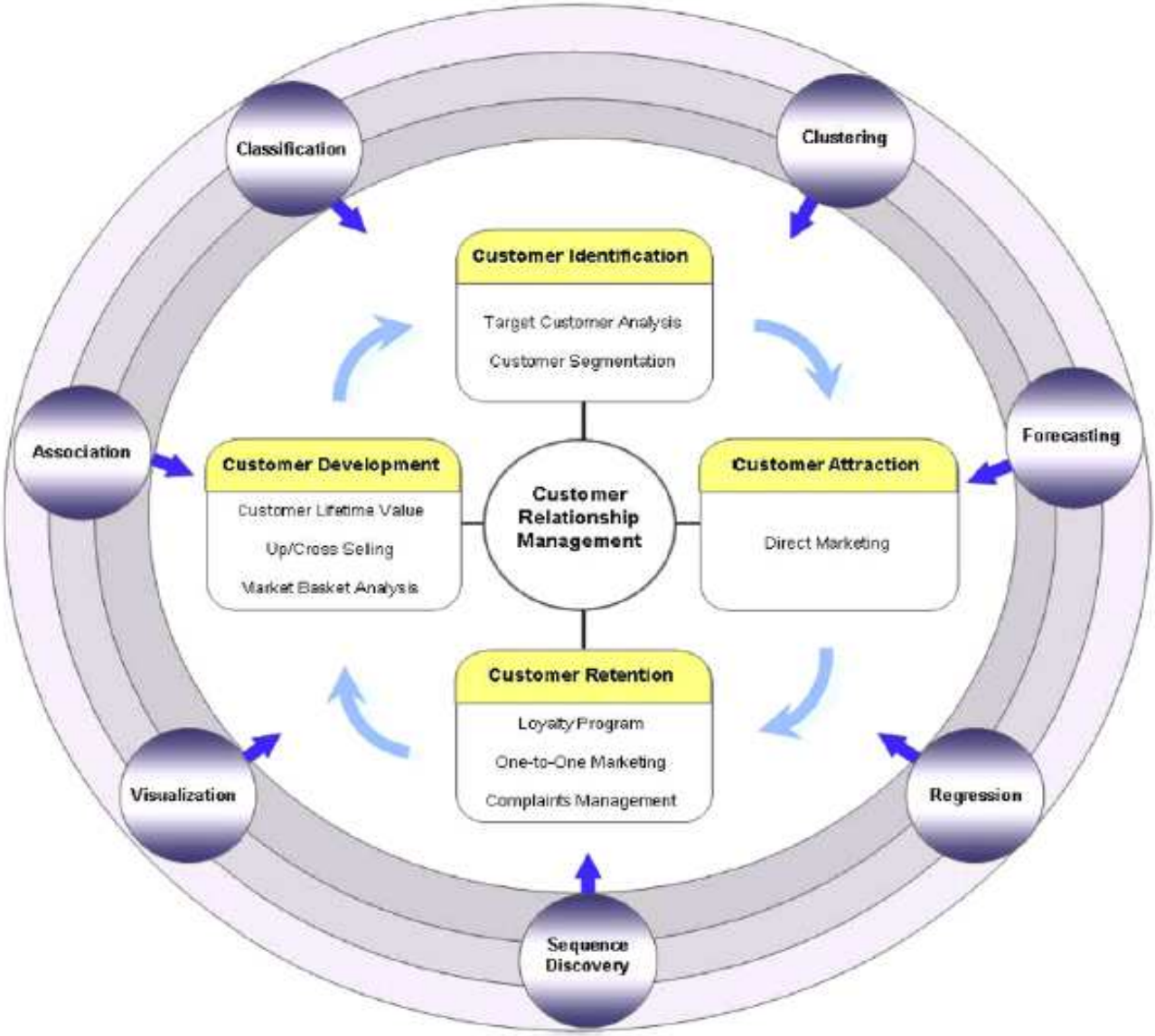
Τα τελευταία χρόνια έχουν γίνει σημαντικές βελτιώσεις, όσον αφορά την ενσωμάτωση της εξόρυξης δεδομένων στη διαδικασία του CRM. Η τάση αυτή αναμένεται να συνεχιστεί, με αποτέλεσμα οι εφαρμογές του CRM να έχουν όλο και περισσότερες δραστηριότητες μάρκετινγκ βασιζόμενες στα αποτελέσματα της εξόρυξης δεδομένων (Thearling 2010). Ενώ οι μελέτες για την εξόρυξη δεδομένων έχουν επικεντρωθεί κυρίως στις τεχνικές, οι μελέτες για τις πελατειακές σχέσεις έχουν επικεντρωθεί στην αλληλεπίδραση του πελάτη και των στρατηγικών της εταιρείας. Η σωστή διαχείριση πελατειακών σχέσεων μπορεί να επιτευχθεί μόνο μέσω της ενοποίησης της διαδικασίας ανακάλυψης γνώσης με τη διαχείριση και χρήση της γνώσης στις στρατηγικές μάρκετινγκ. Αυτό βοηθάει την επιχείριση στην αντιμετώπιση των αναγκών των πελατών, με βάση τις πληροφορίες που γνωρίζει για τον κάθε πελάτη ξεχωριστά, παρά με τη χρήση μίας μαζικής γενίκευσης των χαρακτηριστικών των πελατών (Shaw et al. 2001).

## 3.3 Ταξινόμηση των τεχνικών εξόρυξης δεδομένων στις τέσσερις διαστάσεις του CRM

Η ανάλυση και η κατανόηση των συμπεριφορών και των χαρακτηριστικών των πελατών αποτελούν τα θεμέλια της ανάπτυξης μιας ανταγωνιστικής στρατηγικής CRM

με σκοπό την απόκτηση και τη διατήρηση των δυνητικών πελατών και τη μεγιστοποίηση της αξίας των πελατών. Τα εργαλεία εξόρυξης δεδομένων, τα οποία είναι κατάλληλα στο να εξάγουν και να εντοπίζουν χρήσιμες πληροφορίες και γνώσεις από την τεράστια βάση δεδομένων των πελατών, προσφέρουν μεγάλη υποστήριξη στη λήψη διαφορετικών αποφάσεων CRM (Ngai et al. 2009).

Στην *εικόνα 3.1* απεικονίζεται η ταξινόμηση των τεχνικών εξόρυξης δεδομένων που χρησιμοποιούνται στο CRM, ανάλογα με τη διάσταση του CRM πάνω στην οποία εφαρμόζονται. Οι τεχνικές της εξόρυξης δεδομένων που χρησιμοποιούνται θα πρέπει να βασίζονται στα χαρακτηριστικά των δεδομένων, όπως επίσης και στις απαιτήσεις της επιχείρησης.




Εικόνα 3.1 : Ταξινόμηση των τεχνικών εξόρυξης δεδομένων στο CRM (Ngai et al. 2009).

Πολύ συχνά, απαιτείται να γίνει συνδυασμός των μοντέλων εξόρυξης δεδομένων για την υποστήριξη ή την πρόβλεψη των επιπτώσεων μίας στρατηγικής CRM. Σε μία τέτοια περίπτωση, η ταξινόμηση των μοντέλων εξόρυξης δεδομένων θα βασίζεται στα σημαντικότερα σημεία της στρατηγικής CRM (Ngai et al. 2009). Για παράδειγμα, δεδομένου ότι οι σχέσεις μεταξύ των προϊόντων είναι το κύριο μέλημα της επιχείρησης, στην περίπτωση των προγραμμάτων σταυροειδών πωλήσεων, οι πελάτες μπορούν να κατηγοριοποιηθούν σε ομάδες προτού εφαρμοστεί σε κάθε ομάδα κάποιο μοντέλο συσχετίσεων. Σε τέτοιες περιπτώσεις, το πρόγραμμα σταυροειδών πωλήσεων θα πρέπει να ταξινομηθεί με βάση το μοντέλο συσχετίσεων. Στην περίπτωση του άμεσου μάρκετινγκ, ένα ορισμένο τμήμα των πελατών μπορεί να υποδιαιρεθεί σε ομάδες, έτσι ώστε να σχηματιστούν τα αρχικά τμήματα του μοντέλου τμηματοποίησης. Το πρόγραμμα του άμεσου μάρκετινγκ θα ταξινομηθεί με βάση την τμηματοποίηση, δεδομένου ότι το κύριο μέλημα εδώ είναι η πρόβλεψη της συμπεριφοράς των πελατών.

## 3.4 Προετοιμασία των δεδομένων για τη διαδικασία της εξόρυξης δεδομένων

### 3.4.1 Προετοιμασία των δεδομένων για όλα τα εργαλεία εξόρυξης δεδομένων

Πολύ συχνά, απαιτείται να γίνει προετοιμασία των δεδομένων, προκειμένου να εισέλθουν στην διαδικασία της εξόρυξης δεδομένων. Μάλιστα, μπορεί να δαπανηθεί περισσότερος χρόνος κατά την προετοιμασία των δεδομένων, παρά κατά τη διαδικασία της εξόρυξης δεδομένων. Τα κυριότερα βήματα, προκειμένου τα δεδομένα να αποκτήσουν την απαιτούμενη μορφή, είναι τα εξής (Kimball 1997):

- Η διόρθωση μη συνεπούς μορφής δεδομένων και η διόρθωση μη συμβατής κωδικοποίησης των δεδομένων, συντομογραφιών και σημείων στίξης.

- Η αφαίρεση ανεπιθύμητων ή περιττών πεδίων. Τα δεδομένα μπορεί να περιέχουν πεδία άνευ σημασίας για την ανάλυση που θέλουμε να κάνουμε. Τα εργαλεία της εξόρυξης δεδομένων ενδέχεται να ερμηνεύσουν αυτά τα πεδία ως μετρήσεις ή μεγέθη, ειδικά αν είναι αριθμητικά, και μπορεί να κάνουν κύκλους προσπαθώντας να βρουν συγκεκριμένα μοτίβα με αυτά τα πεδία, ή προσπαθώντας να συσχετίσουν αυτά τα πεδία με πραγματικά δεδομένα.

- Η ερμηνεία των κωδικών σε κείμενο. Η κλασική μορφή «καθαρισμού» των δεδομένων περιλαμβάνει τη βελτίωση ή την αντικατάσταση αινιγματικών κωδικών με ισοδύναμα κειμένου, γραμμένο με αναγνωρίσιμες λέξεις.

- Ο συνδυασμός των δεδομένων που προέρχονται από διάφορες πηγές, όπως τα δεδομένα των πελατών, σε μία κοινή βάση.

- Η εύρεση των πεδίων που έχουν χρησιμοποιηθεί για παραπάνω από έναν σκοπό. Ένας καλός τρόπος για να βρεθούν αυτά τα αρχεία είναι η καταμέτρηση, και ίσως και η δημιουργία μίας λίστας όλων των διαφορετικών τιμών-χρήσεων που υπάρχουν σε ένα πεδίο.

- Ο έλεγχος για τυχόν μη φυσιολογικά, εκτός ορίων ή αδύνατα στοιχεία. Κάποια μετρήσιμα στοιχεία μπορεί να είναι σωστά, αλλά εξαιρετικά ασυνήθιστα. Τέτοιου είδους δεδομένα, είναι προτιμότερο να μαρκαριστούν με μία ειδική σήμανση, έτσι ώστε να μπορούμε να τα συμπεριλάβουμε ή να τα εξαιρέσουμε από την ανάλυσή μας, ανάλογα με την περίπτωση.

- Ο έλεγχος για τυχόν τιμές που λείπουν, ή εάν αυτές έχουν αντικατασταθεί από κάποιον προεπιλεγμένο αριθμό.

- Η εφαρμογή ομοιόμορφης μεταχείρισης σε μηδενικές τιμές. Οι μηδενικές τιμές ενδέχεται να δυσκολέψουν τη λειτουργία του εργαλείου Εξόρυξης Δεδομένων. Σε πολλές περιπτώσεις, η μηδενική αξία αντιπροσωπεύεται από κάποια άλλη ιδιαίτερη τιμή. Πολλές φορές η τιμή -1 θεωρείται ότι αντιπροσωπεύει τις μηδενικές αξίες.

- Η ταξινόμηση μεμονωμένων αρχείων δεδομένων σύμφωνα με ένα από τα συγκεντρωτικά του μεγέθη. Σε ορισμένες περιπτώσεις, μπορεί να είναι επιθυμητός ο εντοπισμός της πώλησης ενός πολύ συγκεκριμένου προϊόντος, όπως ένα ρούχο σε ένα συγκεκριμένο χρώμα και μέγεθος, και από ένα συγκεκριμένο υλικό.

### 3.4.2 Προετοιμασία των δεδομένων ανάλογα με το χρησιμοποιούμενο εργαλείο εξόρυξης δεδομένων

Ανάλογα με το χρησιμοποιούμενο εργαλείο εξόρυξης δεδομένων, μπορεί να χρειαστεί να γίνουν κάποιες επιπλέον μετατροπές στα δεδομένα, πέρα των προηγουμένων, και αυτές είναι οι ακόλουθες:

- Ο διαχωρισμός των πρωτογενών δεδομένων εισόδου σε τρεις ομάδες. Η πρώτη ομάδα δεδομένων χρησιμοποιείται για την κατάρτιση του εργαλείου εξόρυξης δεδομένων. Ένα εργαλείο συσταδοποίησης (clustering tool), ένα εργαλείο νευρωνικών δικτύων (neural network tool) ή ένα εργαλείο δέντρου αποφάσεων (decision tree tool) απορροφά την πρώτη σειρά στοιχείων και ορίζει τις παραμέτρους, από τις οποίες μπορούν να γίνουν οι μελλοντικές ταξινομήσεις και οι προβλέψεις. Το δεύτερο σύνολο δεδομένων, που χρησιμοποιείται στη συνέχεια, ελέγχει αυτές τις παραμέτρους για να δει πόσο καλά αποδίδει το μοντέλο. Όταν το εργαλείο εξόρυξης δεδομένων έχει ρυθμιστεί σωστά στο πρώτο και στο δεύτερο σετ δεδομένων, εφαρμόζεται στη συνέχεια στην τρίτη σειρά αξιολόγησης των δεδομένων, όπου τα συμπλέγματα (clusters), οι ταξινομήσεις και οι προβλέψεις που προέρχονται από το εργαλείο είναι πλήρως αξιόπιστες και μπορούμε να τις χρησιμοποιήσουμε.

- Η προσθήκη υπολογισμένων πεδίων ως εισροές ή ως στόχοι. Για παράδειγμα, ένα υπολογιζόμενο πεδίο, όπως τα κέρδη ή η ικανοποίηση των πελατών, που αντιπροσωπεύει την αξία ενός συνόλου συναλλαγών των πελατών, μπορεί να χρειαστεί να τεθεί ως στόχος για να επιλέξει το εργαλείο εξόρυξης δεδομένων τους πιο κερδοφόρους πελάτες ή για να επιλέξει τη συμπεριφορά που θέλουμε να ενθαρρύνουμε.

- Η διάταξη των συνεχών τιμών σε κλίμακες. Μερικά εργαλεία εξόρυξης δεδομένων, όπως τα δέντρα αποφάσεων, ενθαρρύνουν τη διάταξη αυτή σε διακριτές κλίμακες.

- Η εξομάλυνση των τιμών μεταξύ 0 και 1. Τα εργαλεία νευρωνικών δικτύων συνήθως απαιτούν όλες οι αριθμητικές τιμές να αντιστοιχίζονται σε μια σειρά από το μηδέν έως το ένα.

- Η μετατροπή των κειμένων σε αριθμητικές τιμές. Μερικά εργαλεία εξόρυξης δεδομένων μπορούν να λειτουργήσουν μόνο με αριθμητικά δεδομένα εισόδου. Σε αυτές τις περιπτώσεις, οι διακριτές τιμές κείμενου θα πρέπει να αντικατασταθούν από ειδικούς κωδικούς, όπως για παράδειγμα η αντικατάσταση της περιοχής του κάθε πελάτη με τον αντίστοιχο ταχυδρομικό του κώδικα.

## 3.5 Τα βήματα εφαρμογής της εξόρυξης δεδομένων στη διαχείριση πελατειακών σχέσεων

Σύμφωνα με τον Edelstein (2000), τα βασικά βήματα της εξόρυξης δεδομένων για ένα αποτελεσματικό σύστημα διαχείρισης πελατών είναι τα ακόλουθα:

1. Καθορισμός του προβλήματος της επιχείρησης (define business problem). Κάθε εφαρμογή CRM έχει έναν ή περισσότερους επιχειρηματικούς στόχους για τους οποίους θα πρέπει να οικοδομηθεί το κατάλληλο μοντέλο. Ο αποτελεσματικός εντοπισμός του προβλήματος περιλαμβάνει και έναν τρόπο μέτρησης των αποτελεσμάτων του έργου του CRM.

2. Δημιουργία βάσης δεδομένων μάρκετινγκ (build a marketing database). Είναι απαραίτητη η οικοδόμηση μίας βάσης δεδομένων μάρκετινγκ, διότι οι επιχειρησιακές βάσεις δεδομένων συχνά δεν περιέχουν τα δεδομένα που απαιτούνται, με τη μορφή που απαιτούνται. Επίσης, εάν υπάρχουν ξεχωριστές βάσεις δεδομένων, όπως για παράδειγμα ξεχωριστή βάση δεδομένων για τους πελάτες, ξεχωριστή για τα προϊόντα και ξεχωριστή για τις συναλλαγές, θα πρέπει να γίνει μία ενσωμάτωση όλων αυτών σε μία ενιαία βάση μάρκετινγκ.

3. Εξερεύνηση των δεδομένων (explore the data). Απαραίτητη κρίνεται η σωστή κατανόηση των δεδομένων προκειμένου να οικοδομηθεί ένα αξιόπιστο μοντέλο προβλέψεων. Χρήσιμη θα μπορούσε να είναι η συγκέντρωση κάποιων αριθμητικών δεικτών, όπως για παράδειγμα μέσοι όροι και τυπικές αποκλίσεις, και η εξέταση της διανομής των δεδομένων. Σημαντική βοήθεια προσφέρουν οι γραφικές παραστάσεις και τα εργαλεία απεικόνισης. Η οπτικοποίηση των δεδομένων οδηγεί συχνά σε νέες ιδέες και κατά συνέπεια, στην επιτυχία.

4. Προετοιμασία των δεδομένων για τη μοντελοποίηση ([prepare data for modeling). Αρχικά επιλέγονται οι μεταβλητές πάνω στις οποίες θα κατασκευαστεί το μοντέλο και στη συνέχεια κατασκευάζονται οι νέες μεταβλητές πρόβλεψης, οι οποίες προέρχονται από τα ανεπεξέργαστα δεδομένα. Κατόπιν, επιλέγεται ένα υποσύνολο ή ένα δείγμα των δεδομένων, πάνω στα οποία θα κατασκευαστεί το μοντέλο. Το τελικό στάδιο αφορά τη μετατροπή των μεταβλητών σύμφωνα με τις απαιτήσεις του αλγορίθμου που θα επιλεχθεί για να κατασκευαστεί το μοντέλο.

5. Κατασκευή του μοντέλου εξόρυξης δεδομένων (build model**).** Η κατασκευή του μοντέλου αποτελεί μία επαναληπτική διαδικασία και θα πρέπει να διερευνηθούν εναλλακτικά μοντέλα για να βρεθεί το καταλληλότερο στην επίλυση του προβλήματος της επιχείρησης. Κατά τη διαδικασία αυτή ανίχνευσης του μοντέλου, πολλές φορές μπορεί να οδηγηθούμε σε κάποιο προηγούμενο βήμα και να κάνουμε κάποιες αλλαγές στα δεδομένα, ή ακόμα και να γίνει επαναπροσδιορισμός του αρχικού προβλήματος της επιχείρησης.

6. Αξιολόγηση του μοντέλου (evaluate model). Ίσως το πιο υπερεκτιμημένο μέτρο αξιολόγησης του μοντέλου είναι η ακρίβεια των αποτελεσμάτων. Ένα άλλο μέτρο που χρησιμοποιείται συχνά είναι η ανύψωση (lift), η οποία μετράει τη βελτίωση που επιτυγχάνεται με το μοντέλο πρόβλεψης. Ωστόσο, η μέθοδος αυτή δε λαμβάνει υπόψιν τα κόστη και τα έσοδα, έτσι είναι συχνά προτιμότερο να εξετάσουμε τους δείκτες που αναφέρονται στα κέρδη ή την απόδοση της επένδυσης (ROI).

7. Ανάπτυξη του μοντέλου και αποτελέσματα (deploy model and results). Στην πραγματικότητα, ο τρόπος που ενσωματώνεται η εξόρυξη δεδομένων στην εφαρμογή του CRM καθορίζεται από το είδος της αλληλεπίδρασης της εταιρείας με τον πελάτη. Υπάρχουν δύο βασικοί τρόποι αλληλεπίδρασης: η εισερχόμενη, κατά την οποία οι πελάτες επικοινωνούν με την εταιρεία, και η εξερχόμενη, κατά την οποία η εταιρεία επικοινωνεί με τους πελάτες. Στις εισερχόμενες συναλλαγές, όπως για παράδειγμα μία τηλεφωνική παραγγελία ή παραγγελία μέσω internet, η εφαρμογή θα πρέπει να ανταποκριθεί σε πραγματικό χρόνο, κάτι το οποίο δε συμβαίνει στις εξερχόμενες συναλλαγές.

## 3.6 Η εξόρυξη δεδομένων στον κύκλο ζωής του πελάτη

Για να είναι αποτελεσματικό ένα CRM σύστημα, θα πρέπει να συνδεθούν τα προϊόντα και οι στρατηγικές της εταιρείας με τους στόχους και τους πελάτες της. Ο όρος «κύκλος ζωής του πελάτη» (customer life cycle) αναφέρεται σε όλα τα στάδια της σχέσης μεταξύ της επιχείρησης και του πελάτη. Είναι απολύτως απαραίτητη η κατανόησή του όρου από την εταιρεία, διότι σχετίζεται άμεσα με την κερδοφορία της (Rygielski et al. 2002). Σύμφωνα με τον Edelstein (2000), ο κύκλος ζωής του πελάτη αποτελείται από τρία στάδια, τα οποία είναι: η απόκτηση του πελάτη, η αύξηση της αξίας του πελάτη και η διατήρησή του.

1. Απόκτηση πελατών: Αποτελεί το πρώτο βήμα του CRM και η εξόρυξη δεδομένων μπορεί να συνεισφέρει στη βελτίωση της αποτελεσματικότητας μια εκστρατείας για την απόκτηση πελατών και στην ελαχιστοποίηση του κόστους.

2. Αύξηση της αξίας των υφιστάμενων πελατών:

- ✓ Σταυροειδείς πωλήσεις: Η εξόρυξη δεδομένων, χρησιμοποιώντας τις πληροφορίες των πελατών στις βάσεις δεδομένων, είναι σε θέση να βοηθήσει τον εκπρόσωπο εξυπηρέτησης πελατών να προτείνει τα κατάλληλα επιπρόσθετα προϊόντα στον πελάτη, ή και να μην προτείνει κανένα προϊόν, εάν ο πελάτης δεν είναι δεκτικός σε τέτοιου είδους πωλήσεις. Επίσης, μέσω της εξόρυξης δεδομένων μπορούν να ελαχιστοποιηθούν τα παράπονα των πελατών και να αυξηθεί η κερδοφορία της επιχείρησης.

- ✓ Εξατομίκευση πελατών: Μέσω της συσταδοποίησης της εξόρυξης δεδομένων γίνεται εφικτή η ομαδοποίηση των παρεμφερών προϊόντων, έτσι ώστε κάθε φορά που κάποιος πελάτης δείχνει ενδιαφέρον για ένα προϊόν, η εταιρεία να προβεί στις συστάσεις για αγορά περισσότερων προϊόντων. Με βάση το προφίλ του πελάτη εντοπίζονται οι πελάτες που μπορεί να ενδιαφέρονται για νέα προϊόντα που προστίθενται στον κατάλογο.

3. Διατήρηση των κερδοφόρων πελατών: Για σχεδόν κάθε εταιρεία, το κόστος απόκτησης ενός νέου πελάτη υπερβαίνει το κόστος διατήρησης κερδοφόρων πελατών. Με τη δημιουργία του προφίλ των επικερδών, αλλά και των μη επικερδών πελατών, καθίσταται εφικτότερη η διατήρησή τους ως πελάτες, και ο εντοπισμός των πελατών

που δεν αποφέρουν σημαντικά έσοδα στην επιχείρηση, αλλά θα μπορούσαν να αποφέρουν στο μέλλον.

Ο κύκλος ζωής του πελάτη αποτελεί ένα καλό πλαίσιο για την εφαρμογή της εξόρυξης δεδομένων στο CRM, αφού είναι σε θέση να προβλέψει την κερδοφορία των δυνητικών πελατών, καθώς αυτοί γίνονται ενεργοί, πόσο καιρό θα είναι ενεργοί πελάτες και ποιες είναι οι πιθανότητες να παύσουν να είναι πελάτες. Βέβαια, δε θα είναι ένας ακριβής προγνωστικός δείκτης για το πότε συμβαίνουν οι περισσότερες εκδηλώσεις του κύκλου ζωής, αλλά θα μπορεί να βοηθήσει την επιχείρηση να αναγνωρίζει πρότυπα στα δεδομένα των πελατών της, τα οποία είναι προβλέψιμα. Για παράδειγμα, θα μπορούσε να προβλέψει τη συμπεριφορά που περιβάλλει ένα ιδιαίτερο γεγονός του κύκλου ζωής (π.χ. συνταξιοδότηση) και να βρει άλλους πελάτες σε παρόμοια στάδια ζωής και, κατ' επέκταση, ανάλογες συμπεριφορές.

## 3.7 Marketing Data Intelligence

Το αποτέλεσμα της εφαρμογής της εξόρυξης δεδομένων στον κύκλο ζωής του πελάτη είναι το Marketing Data Intelligence, το οποίο ορίζεται ως «ο συνδυασμός του μάρκετινγκ, με βάση τα δεδομένα, και της τεχνολογίας με σκοπό την μεγιστοποίηση της γνώσης και της κατανόησης των πελατών, και τον συνδυασμό των προϊόντων και δεδομένων συναλλαγής με σκοπό τη βελτίωση της στρατηγικής λήψης αποφάσεων» (Rygielski et al. 2002). Υπάρχουν δύο σημαντικά συστατικά στοιχεία του Marketing Data Intelligence, ο μετασχηματισμός των δεδομένων των πελατών (customer data transformation) και η ανακάλυψη γνώσης των πελατών (customer knowledge discovery).

Τα ακατέργαστα δεδομένα που εξάγονται και μετασχηματίζονται από ένα ευρύ φάσμα εσωτερικών και εξωτερικών βάσεων δεδομένων και η συλλογή όλων αυτών των δεδομένων σε μία κεντρική θέση, όπου μπορούν να αναζητηθούν και να διερευνηθούν, αποτελούν τον μετασχηματισμό των δεδομένων. Η διαδικασία αυτή συνεχίζεται μέσω της ανακάλυψης γνώσης των πελατών, κατά την οποία προέρχονται οι πληροφορίες και μπορούν να εξαχθούν χρήσιμα πρότυπα και συμπεράσματα μέσα από τα δεδομένα. Η διαδικασία αυτή θα πρέπει να παρακολουθείται, έτσι ώστε να διασφαλίζεται ότι τα αποτελέσματα παράγουν αξιοποιήσιμες πληροφορίες (Rygielski et al. 2002).

## 3.8 Τμηματοποίηση πελατών με εργαλεία της εξόρυξης δεδομένων

Οι παραδοσιακές μέθοδοι τμηματοποίησης πελατών στηρίζονται κυρίως σε εμπειρικές ταξινομήσεις ή απλές στατιστικές μεθόδους. Κατηγοριοποιούν τους πελάτες σύμφωνα με κάποιο κοινό συμπεριφορικό χαρακτηριστικό ή ιδιότητα και δεν μπορούν να κάνουν πιο πολύπλοκες αναλύσεις. Επίσης, καθώς η συσσώρευση των δεδομένων γίνεται ολοένα και μεγαλύτερη, οι παραδοσιακές μέθοδοι δεν μπορούν να αντεπεξέλθουν με το μέγεθος της πολυπλοκότητας (Chen et al. 2006).

Πολλές μεγάλες εταιρείες σήμερα κατέχουν terabytes δεδομένων. Η εξόρυξη δεδομένων δίνει τη δυνατότητα στις επιχειρήσεις να εξάγουν όσο το δυνατόν περισσότερες και αξιοποιήσιμες πληροφορίες μέσα από τον τεράστιο, αυτό, όγκο δεδομένων. Πρόκειται, δηλαδή, για μια λύση σε ένα μεγάλο πρόβλημα που αντιμετωπίζουν πολλές επιχειρήσεις: την υπεραφθονία των δεδομένων και μια σχετική έλλειψη κατάλληλου προσωπικού, τεχνολογίας και χρόνου, ώστε να μετατραπούν τα απλά δεδομένα σε ουσιαστικές και αξιοποιήσιμες πληροφορίες σχετικά με τους υπάρχοντες και μελλοντικούς πελάτες (Chen et al. 2006). Οι τεχνικές εξόρυξης δεδομένων που χρησιμοποιούνται σήμερα στην τμηματοποίηση πελατών ανήκουν στην κατηγορία της συσταδοποίησης (clustering) ή των αλγορίθμων των πλησιέστερων γειτόνων (nearest-neighbors algorithms) (Bounsaythip and Runsala 2001).

Η τμηματοποίηση πελατών στο άμεσο μάρκετινγκ έχει καταστεί ιδιαίτερα αποτελεσματική, λόγω της ανάπτυξης των εφαρμογών μάρκετινγκ στις βάσεις δεδομένων. Οι προσεγγίσεις της εξόρυξης δεδομένων παρέχουν αποδοτικούς τρόπους για τον διαχωρισμό των υφιστάμενων πελατών στα τμήματα αλλά και την ανάπτυξη στρατηγικών μάρκετινγκ, προσαρμοσμένες στα συγκεκριμένα τμήματα ή άτομα. Οι τεχνικές μάρκετινγκ στις βάσεις δεδομένων έχουν εξελιχθεί από τα απλά RFM μοντέλα (Recency, Frequency and Monetary), τα οποία περιλαμβάνουν την πρόσφατη πείρα των αγορών των πελατών, τη συχνότητα των αγορών τους, καθώς και το ποσό των χρημάτων που έχουν δαπανήσει με την επιχείρηση, σε στατιστικές τεχνικές όπως η Chisquare τεχνική ανίχνευσης περιοχών ενδιαφέροντος (chisquare automatic interaction detection - CHAID) και το λογιστικό μοντέλο παλινδρόμησης (logistic regression).

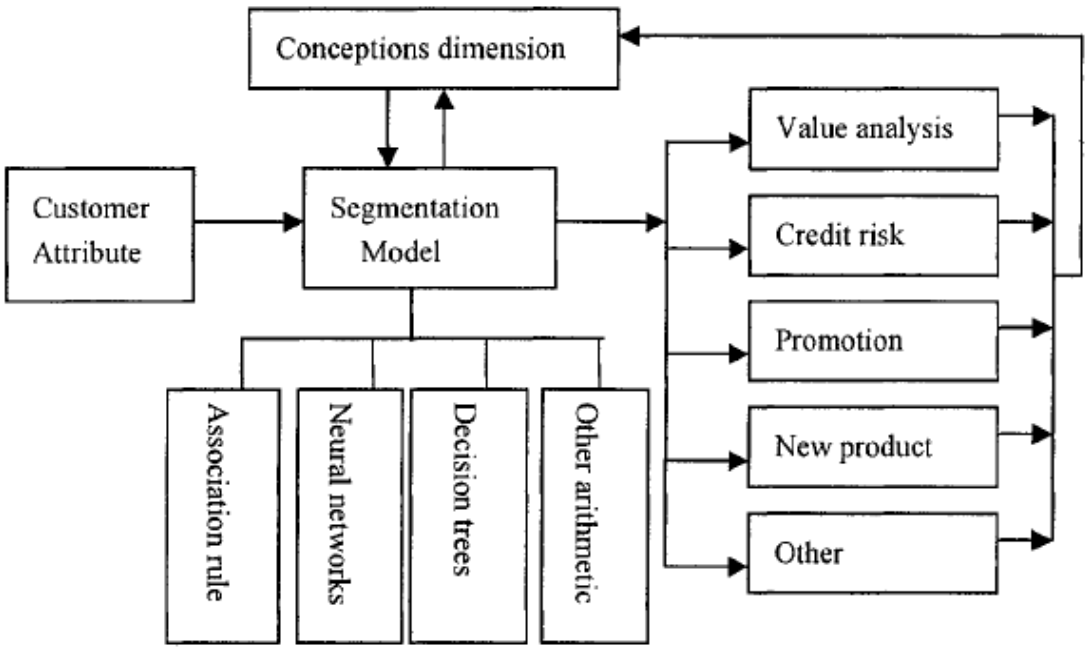
Επίσης, την εμφάνισή τους στις εφαρμογές του μάρκετινγκ στις βάσεις δεδομένων έχουν κάνει και τα νευρωνικά δίκτυα (Yang 2004). Τα εργαλεία της εξόρυξης

δεδομένων παρέχουν ολοκληρωμένη υποστήριξη στις διαδικασίες του management για την απόκτηση και διατήρηση πελατών, την αύξηση της αξίας που προσφέρουν στην επιχείρηση, την πελατειακή ικανοποίηση, καθώς επίσης και την προώθηση της αφοσίωσης των πελατών.

### 3.8.1 Μοντέλο τμηματοποίησης πελατών

Η δημιουργία μιας σχεδιασμένης σχέσης ανάμεσα στα αρχικά χαρακτηριστικά του πελάτη αποτελεί το βασικό βήμα για την πελατειακή τμηματοποίηση, κάνοντας χρήση της εξόρυξης δεδομένων. Τα πελατειακά δεδομένα περιέχουν διανεμημένα και συνεχή χαρακτηριστικά. Θέτοντας το κάθε χαρακτηριστικό του πελάτη ως μια διάσταση (dimension) και θέτοντας κάθε πελάτη ως ένα σωματίδιο (particle), όλοι οι πελάτες μαζί μιας επιχείρησης δημιουργούν έναν πολυδιάστατο χώρο, ο οποίος έχει οριστεί ως «ο χαρακτηριστικός χώρος του κάθε πελάτη» (attribute space of the customer) (Chen et al. 2006).

Η πελατειακή τμηματοποίηση, με χρήση τεχνικών εξόρυξης δεδομένων, είναι δυνατή με τη βοήθεια της λειτουργικής ανάλυσης (functional analysis). Η λειτουργική ανάλυση περιλαμβάνει την ανάλυση της αξίας του πελάτη, την ανάλυση της απόδοσης, την ανάλυση της προώθησης κ.λ.π., βασιζόμενη στα ευρήματα από την σχεδιασμένη σχέση μεταξύ του πελάτη και των βασικών χαρακτηριστικών. Επιπρόσθετα ευρήματα θα ανακαλυφθούν με την ανάπτυξη των πρακτικών management στο CRM. Τα νέα αυτά ευρήματα θα προστεθούν στις θεμελιώδεις διαστάσεις, ενώ η σχεδίαση σχέσεων με τα χαρακτηριστικά των πελατών θα αναδομηθεί. Στην *εικόνα 3.2* απεικονίζεται το μοντέλο τμηματοποίησης πελατών με βάση την εξόρυξη δεδομένων.




Εικόνα 3.2: Μοντέλο τμηματοποίησης πελατών με βάση την εξόρυξη δεδομένων (Chen et al. 2006).

Όπως παρατηρούμε στην *εικόνα 3.2*, η τμηματοποίηση γίνεται με τη βοήθεια τεχνικών εξόρυξης δεδομένων όπως τους κανόνες συσχετίσεων (association rule), τα νευρωνικά δίκτυα (neural network) και τα δέντρα αποφάσεων (decision trees). Το μοντέλο, αρχικά τμηματοποιεί τους πελάτες σύμφωνα με την σχεδιασμένη σχέση και, έπειτα, συνεχίζει τη διαδικασία με τα διάφορα είδη επιχειρησιακών εφαρμογών.

### 3.8.2 Πλεονεκτήματα της μεθόδου τμηματοποίησης με εργαλεία της εξόρυξης δεδομένων

Οι παραδοσιακές μέθοδοι τμηματοποίησης πελατών κατατάσσουν τους πελάτες σε κατηγορίες σύμφωνα με τα απλά χαρακτηριστικά των πελατών ή των αγοραζόμενων προϊόντων, όπως για παράδειγμα η κατηγορία του προϊόντος που αγοράστηκε ή η περιοχή κατοικίας. Επίσης, δεν έχουν τη δυνατότητα να εφαρμόζουν σύνθετες αναλύσεις, όπως τον εντοπισμό των πελατών που έχουν υψηλή δυνητική αξία. Τα μοντέλα εξόρυξης δεδομένων πλεονεκτούν έναντι των παραδοσιακών μεθόδων τμηματοποίησης, προσφέροντας πολλά πλεονεκτήματα, τα οποία είναι τα ακόλουθα (Chen et al. 2006):

1. Βελτίωση των προωθητικών ενεργειών: Το μοντέλο αυτό μπορεί να βοηθήσει την επιχείρηση να ακολουθήσει τις κατάλληλες προωθητικές στρατηγικές, στον κατάλληλο χρόνο και με τα κατάλληλα προϊόντα και υπηρεσίες, με στόχο τους κατάλληλους πελάτες.

2. Ανάλυση της αξίας και της αφοσίωσης των πελατών: Οι δύο αυτές μεταβλητές είναι πολύ σημαντικές γιατί επηρεάζουν τη στρατηγική της επιχείρησης και τις τακτικές διαχείρισης. Οι επιχειρήσεις μέσω του μοντέλου αυτού μπορούν να διαχωρίσουν σε βαθμίδες τους πελάτες σύμφωνα με την αναμενόμενη αξία τους και την αφοσίωσή τους στην εταιρεία.

3. Ανάλυση του πιστωτικού κινδύνου: Η αξιολόγηση του κινδύνου είναι ένας αποτελεσματικός τρόπος για την αξιολόγηση ορισμένων συγκεκριμένων τύπων κινδύνου των πελατών, όπως είναι ο κίνδυνος αθέτησης των υποχρεώσεών τους.

4. Δημιουργία νέων προϊόντων έρευνας και ανάπτυξης: Οι επιχειρήσεις μπορούν να ανακαλύψουν τις προτιμήσεις των πελατών τους μέσω της ανάλυσης πελατών με βάση την εξόρυξη δεδομένων, και να σιγουρευτούν ότι θα υπάρξει συγκεκριμένη ζήτηση για το σχεδιαζόμενο προϊόν.

5. Η επιβεβαίωση της αγοράς-στόχου: Η τμηματοποίηση πελατών με βάση την εξόρυξη δεδομένων είναι σε θέση να εντοπίσει ρητά την αγορά στην οποία απευθύνεται το προϊόν-υπηρεσία της επιχείρησης και να καταστήσει σαφείς τις στοχευμένες πελατειακές ομάδες.

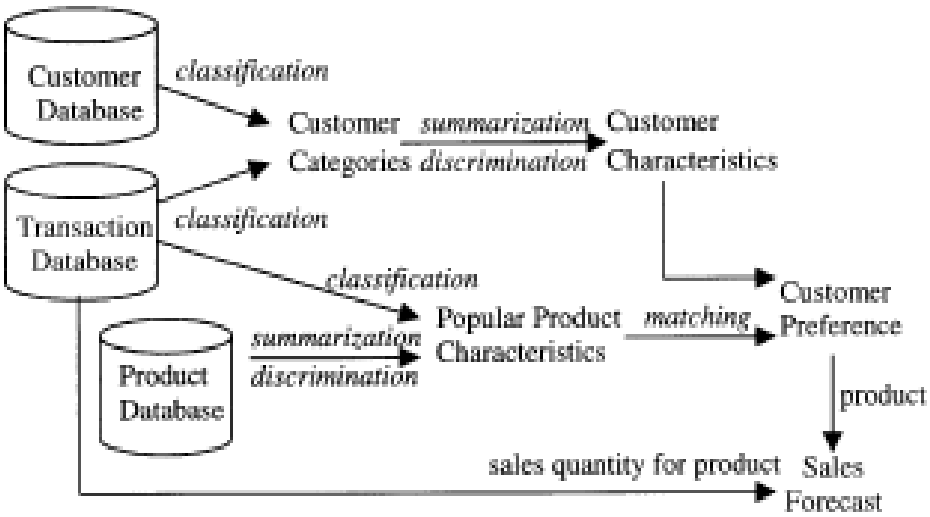
## 3.9 Δημιουργία προφίλ πελατών με εργαλεία της εξόρυξης δεδομένων

Η δημιουργία προφίλ πελατών είναι ένας τρόπος εφαρμογής εξωτερικών δεδομένων σε ένα πληθυσμό πιθανών πελατών. Ανάλογα με τα διαθέσιμα στοιχεία, αυτά μπορούν να χρησιμοποιηθούν για να προσελκυστούν νέοι πελάτες ή να εγκαταλειφθούν υπάρχοντες «κακοί» πελάτες, αυτοί δηλαδή που δεν πραγματοποιούν επαναλαμβανόμενες αγορές και συνεπώς δεν αποφέρουν σημαντικά έσοδα στην επιχείρηση. Ο στόχος είναι να προβλεφθούν συμπεριφορές με βάση τις πληροφορίες που υπάρχουν για κάθε πελάτη.

Ο μακροπρόθεσμος στόχος της δημιουργίας προφίλ του πελάτη είναι να μετατρέψει την κατανόηση των χαρακτηριστικών της πελατειακής βάσης σε μια αυτοματοποιημένη αλληλεπίδραση με αυτούς (Thearling 2000).

Για το λόγο αυτό, οι ανάγκες της σημερινής αγοράς απαιτούν ένα ευρύ φάσμα διαδικασιών και πληροφοριακών εργαλείων. Αυτά τα εργαλεία χρησιμοποιούνται στη συλλογή δεδομένων και την απλοποίηση της διαδικασίας εξόρυξης γνώσης. Τα εργαλεία εξόρυξης δεδομένων χρησιμοποιούνται για να εντοπίσουν ομάδες με κοινά χαρακτηριστικά στα ιστορικά δεδομένα, όπως για παράδειγμα κριτήρια επιλογής για λίστες αλληλογραφίας ή για να εντοπιστούν αγορές με υψηλές δυνατότητες επέκτασης. Εν συντομία, τα εργαλεία εξόρυξης δεδομένων είναι σε θέση να εντοπίσουν, μέσα από τα δεδομένα, ανθρώπινα ερμηνεύσιμα μοτίβα, όπως επίσης και να χρησιμοποιούν κάποιες μεταβλητές για να προβλέψουν άγνωστες ή μελλοντικές αξίες άλλων μεταβλητών (Bounsaythip and Runsala 2001).

Στην *εικόνα 3.3* απεικονίζεται ένα σύστημα δημιουργίας του προφίλ πελατών, το οποίο χρησιμοποιεί τεχνικές εξόρυξης δεδομένων.




Εικόνα 3.3: Σύστημα δημιουργίας προφίλ πελατών (Shaw et al. 2001).

## 3.10 Αξιολόγηση της αφοσίωσης των πελατών

Η δημιουργία μιας πιστής πελατειακής βάσης (customer loyalty) δεν αφορά μόνο στη διατήρηση του αριθμού των πελατών με την πάροδο του χρόνου, αλλά και στην ενίσχυση της σχέσης της εταιρείας με τους πελάτες, η οποία θα ενθαρρύνει τις μελλοντικές αγορές τους και θα αυξήσει το επίπεδο αφοσίωσης και υποστήριξης. Γνωρίζοντας το επίπεδο αφοσίωσης των πελατών της, η επιχείρηση είναι σε θέση να κατανοήσει το πώς οι προσπάθειές της για τη διατήρηση καλών σχέσεων μπορούν να συμβάλουν στο επίπεδο των κερδών της (Hosseini et al. 2010). Κατά τον Payne (2002), το κόστος για την απόκτηση ενός νέου πελάτη είναι πέντε φορές μεγαλύτερο από το κόστος διατήρησης ενός παλιού πελάτη. Σύμφωνα με τους Eriksson και Vaghult (2000), υπάρχουν πολλές μεταβλητές που επηρεάζουν το επίπεδο πιστότητας των πελατών και, κατά συνέπεια, την παραμονή τους ως πελάτες. Οι μεταβλητές αυτές είναι η ποιότητα της σχέσης, η εμπιστοσύνη, το ποσοστό συμμετοχής, η ικανοποίηση, η εξέλιξη της αγοράς, οι οργανωτικές αλλαγές και το κόστος αλλαγής προμηθευτή.

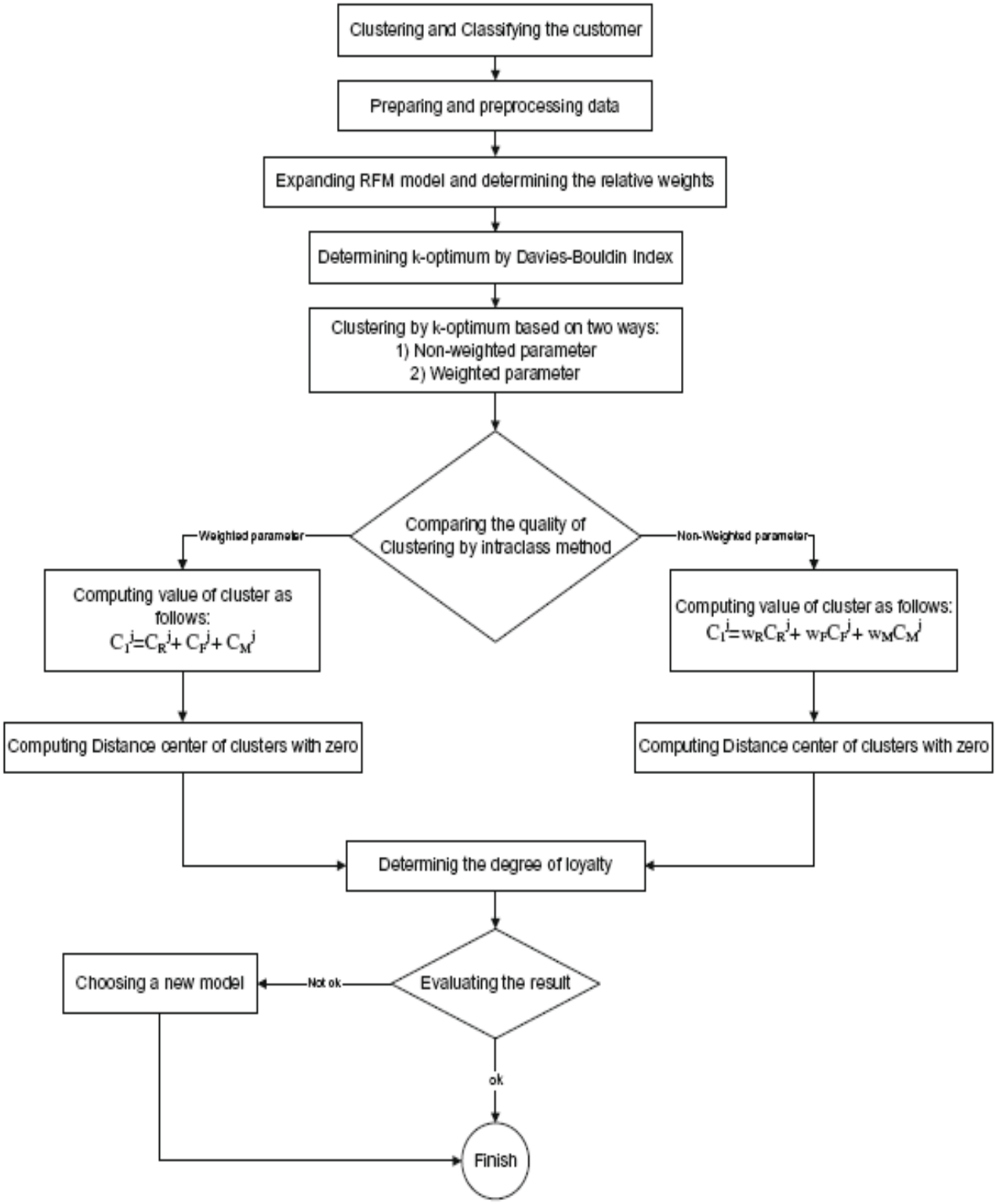
Οι Hosseini, Maleki και Gholamian στο άρθρο τους (2010), χρησιμοποιώντας την RFM ανάλυση και τον αλγόριθμο k-means της εξόρυξης δεδομένων, προτείνουν μία συγκεκριμένη διαδικασία για την ταξινόμηση της αξίας των πελατών, με βάση τα προϊόντα που αγοράζουν. Σύμφωνα με το μοντέλο αυτό, αξιολογείται ο βαθμός αφοσίωσης στο προϊόν, έτσι ώστε να αξιοποιηθεί καταλληλότερα το εφαρμοζόμενο CRM και να μεγιστοποιηθούν τα κέρδη. Στην *εικόνα 3.4* απεικονίζεται η μέθοδος αυτή βήμα προς βήμα.

Η RFM τεχνική εντοπίζει τη συμπεριφορά του πελάτη και αναπαριστά τα χαρακτηριστικά της συμπεριφοράς του με τρεις μεταβλητές ως εξής (Sohrabi and Khanlari 2007):

- Το πρόσφατο διάστημα από την τελευταία αγορά, το οποίο αναφέρεται στο διάστημα από τη στιγμή που εκδηλώθηκε η τελευταία καταναλωτική συμπεριφορά μέχρι και σήμερα.

- Η συχνότητα των αγορών, η οποία αναφέρεται στον αριθμό των συναλλαγών σε μία συγκεκριμένη περίοδο.

- Χρηματική αξία της αγοράς, η οποία αναφέρεται στο ποσό των χρημάτων που δαπάνησε ο πελάτης σε μία συγκεκριμένη περίοδο.




Εικόνα 3.4: Μέθοδος αξιολόγησης της αφοσίωσης των πελατών (Hosseini et al. 2010).

Όπως φαίνεται και στην *εικόνα 3.4*, η διαδικασία αξιολόγησης της αφοσίωσης των πελατών, με τη χρήση των εργαλείων της εξόρυξης δεδομένων, είναι χωρισμένη στα εξής έξι βήματα (Hosseini et al. 2010):

1. Επιλογή του συνόλου των δεδομένων και επεξεργασία των δεδομένων.

2. Καθορισμός της πρόσφατης περιόδου, της συχνότητας και της χρηματικής αξίας των αγορών, όπως επίσης και της διάρκειας δραστηριότητας του προϊόντος, ως τις παραμέτρους του μοντέλου. Στη συνέχεια καθορίζεται το βάρος της κάθε παραμέτρου και αποδίδεται η κλίμακα των RMF χαρακτηριστικών για το σκοπό της ομαδοποίησης (clustering). Στη συνέχεια διαχωρίζονται όλα τα τμήματα της εταιρείας σε 34 κατηγορίες, ενώ αξιολογείται η βέλτιστη τιμή K, ώστε να μπορέσει να χρησιμοποιηθεί αργότερα στον αλγόριθμο k-means. Έπειτα, τα τμήματα συγκεντρώνονται με δύο τρόπους, οι οποίοι βασίζονται σε σταθμισμένες και μη-σταθμισμένες παραμέτρους, προκειμένου να συγκριθεί η ποιότητα των τμημάτων.

3. Υπολογισμός του ολοκληρωμένου ποσοστού του κάθε τμήματος (cluster), το οποίο συμβολίζουμε με F, το οποίο είναι: $C_1^j = W_R\, C_R^j + W_F\, C_F^j + W_M\, C_M^j$, όπου $W_R$, $W_F$, και $W_M$ είναι οι σχετικής σημασίας RMF μεταβλητές.

4. Υπολογισμός της απόστασης μεταξύ του μέσου του κάθε τμήματος με το σημείο μηδέν, και την απόσταση αυτή τη συμβολίζουμε με D.

5. Επίτευξη της κατάλληλης εξίσωσης για την αξιολόγηση της CLV κατάταξης με βάση τις δύο παραμέτρους, την D και την F.

6. Τέλος, διαχωρισμός του συνόλου των δεδομένων σε δοκιμαστικά δεδομένα και επικύρωση των στοιχείων για την αξιολόγηση των αποτελεσμάτων.

## 3.11 Ανίχνευση ακραίων τιμών (outlier detection)

Σε αντίθεση με τις παραδοσιακές τεχνικές της εξόρυξης δεδομένων, οι οποίες έχουν ως στόχο τον εντοπισμό του γενικού πλαισίου που ισχύει για την πλειοψηφία των δεδομένων, η ανίχνευση ακραίων τιμών στοχεύει στον εντοπισμό των σπάνιων

στοιχείων, των οποίων οι συμπεριφορές είναι πολύ ιδιαίτερες, σε σύγκριση με τη μεγαλύτερη ποσότητα των δεδομένων. Μία ακραία τιμή είναι μία παρατήρηση, η οποία αποκλίνει τόσο πολύ από τις υπόλοιπες παρατηρήσεις, ώστε να προκαλεί υποψίες ότι δημιουργήθηκε από έναν διαφορετικό μηχανισμό.

Η μελέτη της ιδιαίτερης συμπεριφοράς των ακραίων τιμών συνεισφέρει στην ανακάλυψη πολύτιμης γνώσης και επιτρέπει στους φορείς λήψης αποφάσεων της επιχείρησης να πραγματοποιήσουν κέρδη ή να βελτιώσουν την ποιότητα των παρεχόμενων υπηρεσιών. Έτσι, η εξόρυξη ακραίων τιμών αποτελεί μία σημαντική έρευνα εξόρυξης δεδομένων με πολλές εφαρμογές σε διάφορους τομείς, όπως η ανίχνευση απάτης μέσω πιστωτικών καρτών, η ανακάλυψη εγκληματικών δραστηριοτήτων στο ηλεκτρονικό εμπόριο, το μάρκετινγκ, και άλλα (He et al. 2004).

# ΚΕΦΑΛΑΙΟ 4: ΤΟ WEB MINING ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ

## 4.1 Εισαγωγή

Με την ανάπτυξη του ηλεκτρονικού εμπορίου και των διαθέσιμων πηγών πληροφόρησης στο διαδίκτυο, γίνεται ολοένα και περισσότερο απαραίτητη στους χρήστες η χρήση αυτοματοποιημένων εργαλείων για τον εντοπισμό των επιθυμητών πηγών πληροφόρησης. Αυτοί οι παράγοντες συνετέλεσαν στην ανάγκη για τη δημιουργία ευφυών συστημάτων, τα οποία μπορούν να εξάγουν αποτελεσματικά χρήσιμες πληροφορίες. Για το λόγο αυτό, δημιουργήθηκε το web mining, το οποίο έχει ως στόχο τη μελέτη των εφαρμογών της εξόρυξης δεδομένων στο διαδίκτυο. Στο κεφάλαιο αυτό παρουσιάζονται οι κατηγορίες και η δομή του web mining και αναλύονται τα εργαλεία ανακάλυψης προτύπων που χρησιμοποιεί. Στη συνέχεια, παρουσιάζεται η εφαρμογή του στο ηλεκτρονικό εμπόριο και τη διαχείριση πελατειακών σχέσεων.

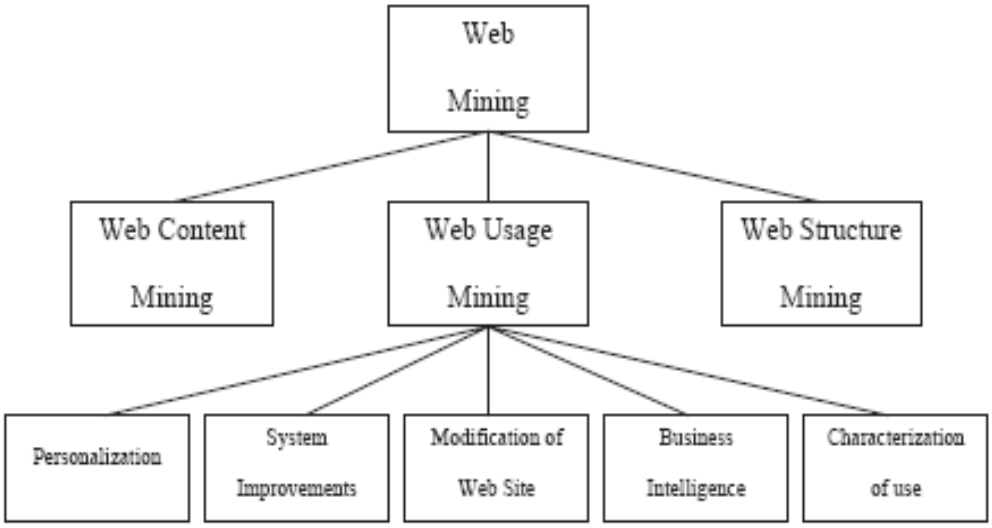
## 4.2 Έννοιες - Ορισμοί

Όταν η οικονομία εξελίσσεται και από προσανατολισμένη στο προϊόν γίνεται πελατοκεντρική, η προώθηση των προϊόντων και υπηρεσιών επικεντρώνονται στην προσαρμοστικότητα και την εξατομίκευση, και η διαχείριση πελατειακών σχέσεων καθίσταται το επίκεντρο ενδιαφέροντος της επιχείρησης. Το διαδίκτυο αποτελεί την καλύτερη βάση για την εφαρμογή της εξόρυξης δεδομένων, η οποία είναι σε θέση να βοηθήσει τον διαχειριστή της ιστοσελίδας να προσδιορίσει το προφίλ των πελατών, να αναλύσει τις συμπεριφορές τους και να έχει άμεση ανατροφοδότηση, ώστε να δημοσιεύονται εξατομικευμένες ηλεκτρονικές πληροφορίες και εν τέλει, να παρέχεται η καλύτερη δυνατή διαχείριση των πελατειακών σχέσεων.

Το web mining ορίζεται γενικά ως η ανακάλυψη και ανάλυση χρήσιμων πληροφοριών από το διαδίκτυο (Cooley et al. 1997), ή πιο απλά, είναι η εφαρμογή των τεχνικών της εξόρυξης δεδομένων στις μεγάλες αποθήκες δεδομένων του διαδικτύου (Mobasher et al. 1996). Αυτός ο ορισμός περιγράφει την αυτόματη αναζήτηση πηγών πληροφοριών στο διαδίκτυο, όπως είναι για παράδειγμα το web content mining, και την ανακάλυψη μοτίβων μέσω της πρόσβασης των χρηστών στους διακομιστές, όπως για παράδειγμα το web usage mining.

## 4.3 Οι κατηγορίες του Web Mining

Σύμφωνα με τους Cooley, Mobasher και Srivastava (1997), το web mining χωρίζεται σε δύο επιμέρους κατηγορίες, το web content mining και το web usage mining, ενώ οι περισσότεροι συγγραφείς (Huang et al. 2003, Araya et al. 2004, Purandare 2008) αναφέρονται σε ακόμα μία κατηγορία, το web structure mining. Στην *εικόνα 4.1* απεικονίζονται οι επιμέρους κατηγορίες του web mining. Η κατηγοριοποίηση αυτή δίνει περισσότερη έμφαση στις περιοχές εφαρμογής του web usage mining, το οποίο χρησιμοποιείται περισσότερο για τους σκοπούς της διαχείρισης πελατειακών σχέσεων.




Εικόνα 4.1: Οι κατηγορίες του web mining (Araya et al. 2004).

### 4.3.1 Web Content Mining

Το web content mining είναι η ανακάλυψη χρήσιμων πληροφοριών από τα περιεχόμενα των ιστοσελίδων (Huang et al. 2003), και μπορεί να προσεγγιστεί και ταξινομηθεί υπό δύο διαφορετικές σκοπιές, την προσέγγιση με βάση τον πράκτορα (agent-based approach) και την προσέγγιση της βάσης δεδομένων (database approach) (Cooley et al. 1997). Σύμφωνα με την πρώτη προσέγγιση, τα συστήματα web mining μπορούν να ταξινομηθούν σε τρεις επιμέρους κατηγορίες:

1. Ευφυείς πράκτορες αναζήτησης (intelligent search agents): πραγματοποιούν αναζήτηση πληροφοριών και, με βάση τα κύρια χαρακτηριστικά και προφίλ των πελατών, οργανώνουν και ερμηνεύουν τις πληροφορίες αυτές.
2. Φιλτράρισμα/κατηγοριοποίηση των πληροφοριών (information filtering/ categorization): χρησιμοποιούν ποικίλες τεχνικές ανάκτησης πληροφοριών από αρχεία εγγράφων στο διαδίκτυο, τις οποίες φιλτράρουν αυτόματα και κατηγοριοποιούν.
3. Εξατομικευμένοι διαδικτυακοί πράκτορες (personalized web agents): μαθαίνουν τις προτιμήσεις των χρηστών και ανακαλύπτουν διαδικτυακές πηγές πληροφοριών, βασιζόμενοι σε αυτές τις προτιμήσεις.

### 4.3.2 Web Usage Mining

Το web usage mining είναι η αυτόματη ανακάλυψη πληροφοριών από τις συμπεριφορές των χρηστών στις ιστοσελίδες του διαδικτύου (Huang et al. 2003). Οι επιχειρήσεις, κατά τις καθημερινές τους λειτουργίες, συλλέγουν μεγάλο όγκο πληροφοριών, ο οποίος παράγεται αυτόματα από τους διακομιστές του διαδικτύου και συλλέγεται και καταγράφεται στα αρχεία των διακομιστών. Η ανάλυση των δεδομένων αυτών μπορεί να βοηθήσει τους οργανισμούς να προσδιορίζουν την αξία των πελατών, τις στρατηγικές μάρκετινγκ των προϊόντων, την αποτελεσματικότητα των προωθητικών ενεργειών και άλλα. Επίσης, μπορεί να προέλθουν πληροφορίες σχετικά με την αναδιάρθρωση της ιστοσελίδας, ώστε να δημιουργηθεί μία δυναμικότερη παρουσία στον ιστοχώρο και να γίνει αποτελεσματικότερη η επικοινωνία των ομάδων εργασίας και της οργανωτικής δομής. Όσον αφορά τις διαφημίσεις προϊόντων στο διαδίκτυο, μέσω της ανάλυσης των πληροφοριών των χρηστών, επιτυγχάνεται η καλύτερη στόχευση των διαφημιστικών μηνυμάτων σε συγκεκριμένες ομάδες χρηστών. Τα περισσότερα υπάρχοντα εργαλεία ανάλυσης στο διαδίκτυο προβλέπουν μηχανισμούς αναφοράς των δραστηριοτήτων των χρηστών, αλλά και διάφορες μορφές φιλτραρίσματος των δεδομένων (Cooley et al. 1997).

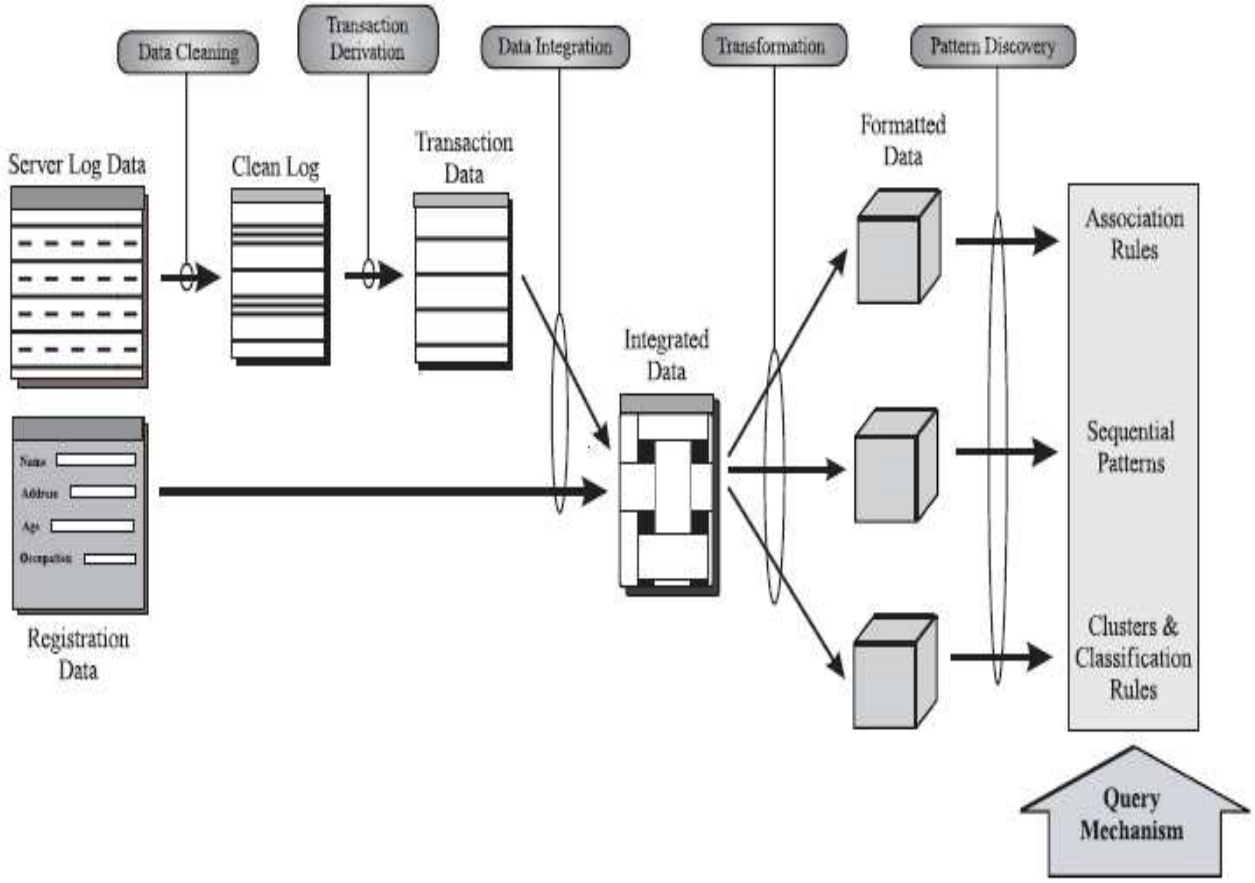
### 4.3.3 Web Structure Mining

Το web structure mining έχει ως στόχο τη δημιουργία μίας σύνοψης της διάρθρωσης των ιστοσελίδων στο διαδίκτυο (Madria et al. 1999). Συγκεκριμένα, εστιάζει την προσοχή του στις πληροφορίες που περιέχουν οι ιστοσελίδες και είναι σε θέση να εντοπίζει τις ομοιότητες μεταξύ των σελίδων ή να βρίσκει τις σελίδες που περιέχουν ένα συγκεκριμένο θέμα (Purandare 2008).

## 4.4 Η δομή του Web Mining

Η προτεινόμενη δομή των Mobasher, Jain, Han και Srivastava (1996) διαχωρίζει τη διαδικασία του web mining σε δύο κύρια μέρη. Το πρώτο μέρος περιλαμβάνει τις διαδικασίες μετατροπής των δεδομένων σε κατάλληλες επεξεργάσιμες μορφές, και το

δεύτερο περιλαμβάνει τις εφαρμογές των τεχνικών εξόρυξης δεδομένων. Η δομή του web mining, απεικονίζεται στην *εικόνα 4.2*, και το μοντέλο ονομάζεται Webminer.




Εικόνα 4.2: Η δομή του Webminer (Mobasher et al. 1996).

Όταν ένας χρήστης επισκέπτεται μία ιστοσελίδα, αυτόματα οι διαχειριστές της σελίδας αποκτούν πρόσβαση σε πολλά δεδομένα, όπως αρχεία εικόνων, ήχων και βίντεο, συντεταγμένες περιοχών όπου ο χρήστης μπορεί να «κλικ άρει», εκτελέσιμα αρχεία και αρχεία μορφής HTML. Έτσι, τα αρχεία καταγραφής του διακομιστή περιέχουν πολλές πληροφορίες, οι οποίες μπορεί να είναι περιττές ή και άσχετες με τις λειτουργίες της εξόρυξης δεδομένων. Η διαδικασία κατά την οποία απομακρύνονται όλα τα περιττά και άσχετα δεδομένα από τον διακομιστή ονομάζεται καθαρισμός των δεδομένων (data cleaning). Όπως φαίνεται και στην *εικόνα 4.2*, η δεύτερη πηγή πληροφοριών, πέρα από τα προσβάσιμα αρχεία του χρήστη, αποτελούν τα δεδομένα που εισέρχονται με την εγγραφή του χρήστη στην ιστοσελίδα, τα οποία θα πρέπει να ενσωματωθούν με τα προηγούμενα δεδομένα.

Μετά τον καθαρισμό των δεδομένων, οι καταχωρήσεις στα αρχεία καταγραφής πρέπει να κατανεμηθούν σε λογικές ομάδες, οι οποίες αντιπροσωπεύουν τις συναλλαγές μεμονωμένων χρηστών και έπειτα, διαμορφώνονται κατάλληλα για να συνάδουν με το μοντέλο δεδομένων της εφαρμογής εξόρυξης δεδομένων. Τέλος, ένας μηχανισμός (query mechanism) επιτρέπει στον αναλυτή να έχει μεγαλύτερο έλεγχο της διαδικασίας, καθορίζοντας διάφορους περιορισμούς.

Η ανακάλυψη προτύπων-μοτίβων μέσω της διαδικασίας του webminer δε θα ήταν ιδιαίτερα χρήσιμη εάν δεν υπήρχαν οι μηχανισμοί και τα εργαλεία ανάλυσης των αποτελεσμάτων, αφού τα αποτελέσματα των αλγορίθμων εξόρυξης γνώσης τις περισσότερες φορές δεν είναι σε μορφή κατανοητή από τους ανθρώπους. Αυτές οι τεχνικές αντλούνται από διάφορους τομείς, όπως η στατιστική, η γραφική απεικόνιση, η οπτικοποίηση των δεδομένων, η ανάλυση χρηστικότητας (usability analysis) και η αναζήτηση σε βάσεις δεδομένων (database querying) (Cooley et al. 1997).

## 4.5 Πηγές διαδικτυακών δεδομένων (web data)

Στο διαδίκτυο απαιτούνται ποικίλες μορφές δεδομένων. Πολλοί ερευνητές ταξινομούν τα δεδομένα σύμφωνα με τα ιδιαίτερα χαρακτηριστικά τους. Για παράδειγμα, τα δεδομένα μπορούν να ταξινομηθούν σε δεδομένα επιπέδου διακομιστή (server level), επιπέδου πελάτη (client level) και επιπέδου μεσολάβησης (proxy level), ανάλογα με την τοποθεσία από την οποία συλλέγονται (Srivastava et al. 2000). Τα δεδομένα από το επίπεδο διακομιστή ονομάζονται server logs, ενώ τα δεδομένα από το επίπεδο πελάτη είναι γνωστά ως πληροφορίες δραστηριότητας του χρήστη (client-side data) και δημιουργούνται στον browser του πελάτη. Τα δεδομένα από το επίπεδο μεσολάβησης περιλαμβάνουν πληροφορίες για τις δραστηριότητες του χρήστη και βρίσκονται στον διακομιστή μεσολάβησης (proxy server).

Οι Ahn, Kim και Han στο άρθρο τους (2003) ταξινομούν τα διαδικτυακά δεδομένα, από τη σκοπιά του μάρκετινγκ, ως εξής:

- Log data: Είναι οι πληροφορίες που προέρχονται από τα συστήματα web server. Τα δεδομένα αυτά περιέχουν πληροφορίες σχετικά με τα προσβάσιμα και αναφερόμενα δεδομένα των χρηστών.

- User data: Είναι οι πληροφορίες που συλλέγονται από τον πελάτη και δεν μπορούν να προέλθουν από τα log data. Τα δεδομένα αυτά περιλαμβάνουν διάφορες μορφές πληροφοριών, όπως το φύλο, η ηλικία, η διεύθυνση ηλεκτρονικού ταχυδρομείου, τον τρόπο πρόσβασης στην ιστοσελίδα και άλλα.

- Market data: Περιλαμβάνουν πληροφορίες σχετικά με τα προϊόντα που παράγονται στην εταιρεία και το ηλεκτρονικό εμπόριο. Τα δεδομένα αυτά χρησιμοποιούνται για την διαμόρφωση στρατηγικών, όπως στρατηγικές πώλησης και διαχείρισης πελατών, όπως επίσης και για την ανάλυση σταυροειδών πωλήσεων, την ανάλυση καλαθιού μάρκετινγκ και την ανάλυση αξίας των πελατών.

## 4.6 Εργαλεία ανακάλυψης προτύπων-μοτίβων του web mining

Τα αναδυόμενα εργαλεία για την ανακάλυψη μοτίβων στα δεδομένα των χρηστών κάνουν χρήση εξελιγμένων τεχνικών τεχνητής νοημοσύνης, εξόρυξης δεδομένων και θεωρίες της ψυχολογίας, προκειμένου να εξάγουν γνώση από τα δεδομένα που έχουν συλλεχθεί (Cooley et al. 1997). Τα εργαλεία που χρησιμοποιούνται στο web mining είναι τα εξής (Huang et al. 2003):

1. Η περιγραφική/περιληπτική παρουσίαση (description/summarization): Στο web mining, η περιγραφική παρουσίαση επιτρέπει τη δημιουργία μίας σύνοψης των χαρακτηριστικών των πελατών που ξοδεύουν ένα συγκεκριμένο ποσό το χρόνο.

2. Η κατάταξη (classification): Επιτρέπει την ανάπτυξη προφίλ των πελατών με βάση τις διαθέσιμες δημογραφικές πληροφορίες ή ανάλογα με το που «κλικ άρουν» οι χρήστες.

3. Η εκτίμηση/πρόβλεψη (estimation/prediction): Μπορεί να εφαρμοστεί σε συνδυασμό με τις τεχνικές OLAP (OnLine Analytical Processing) προκειμένου να γενικευτούν οι ιδιότητες των ομάδων των ατόμων που επισκέπτονται μία τοποθεσία στο διαδίκτυο.

4. Οι κανόνες συσχέτισης (association rules): Προσδιορίζουν τα προϊόντα που έχουν την μεγαλύτερη πιθανότητα να αγοραστούν ή να προβληθούν από τους χρήστες σε μία περίοδο. Αν αυτά τα στοιχεία δεν τοποθετηθούν μαζί έναν ηλεκτρονικό κατάλογο, υπάρχουν πιθανότητες ο πελάτης να ξεχάσει να επιλέξει αυτά τα αντικείμενα προς αγορά.

5. Η συσταδοποίηση (clustering): Μέσω της συσταδοποίησης των πληροφοριών των χρηστών που επισκέπτονται την ιστοσελίδα, διευκολύνεται η ανάπτυξη και εκτέλεση των μελλοντικών στρατηγικών μάρκετινγκ της εταιρείας.

Όπως έχουμε αναφέρει, η εξόρυξη γνώσης από τα δεδομένα των χρηστών του διαδικτύου έχει τη δυνατότητα να ανακαλύπτει πολύτιμες πληροφορίες. Βέβαια, το συνολικό έργο του web mining δεν είναι η απλή εφαρμογή των αλγορίθμων εξόρυξης δεδομένων στα νέα δεδομένα. Λόγω των πολλών ιδιαίτερων χαρακτηριστικών του μοντέλου πελάτη – διακομιστή στο διαδίκτυο, συμπεριλαμβανομένων των ριζικών διαφορών μεταξύ των ομαδοποιήσεων των δεδομένων στις αποθήκες δεδομένων του ιστοχώρου, είναι αναγκαίο να αναπτυχθεί ένα νέο πλαίσιο προκειμένου να καταστεί δυνατή η εφαρμογή της εξόρυξης γνώσης. Συγκεκριμένα, προτού εκτελεστούν οι αλγόριθμοι εξόρυξης δεδομένων στα δεδομένα, θα πρέπει πρωτίστως να γίνει κάποια προετοιμασία και επεξεργασία των δεδομένων, όπως για παράδειγμα η ανάπτυξη τεχνικών καθαρισμού/φιλτραρίσματος των δεδομένων, ώστε να εξαλειφθούν οι ακραίες τιμές ή τα άσχετα δεδομένα, η ομαδοποίηση μεμονωμένων προσβάσεων στις σελίδες σε ενότητες, και η χρησιμοποίηση ειδικευμένων αλγορίθμων εξόρυξης δεδομένων (Mobasher et al. 1996).

## 4.7 Το Web Mining στο ηλεκτρονικό εμπόριο

Είναι προφανές ότι, στο εικονικό περιβάλλον, οι πληροφορίες για τα προϊόντα δεν είναι μόνο άμεσα διαθέσιμες, αλλά και ο όγκος τους είναι τεράστιος. Ωστόσο, σε αντίθεση με το φυσικό περιβάλλον αγορών, όπου οι καταναλωτές έχουν την ευκαιρία να ενημερωθούν για το προϊόν πρόσωπο με πρόσωπο, το εικονικό περιβάλλον δεν επιτρέπει τέτοιου είδους συναλλαγές. Ως εκ τούτου, οι καταναλωτές πρέπει να έχουν ιδιαίτερες δεξιότητες και γνώσεις για να αναζητήσουν αποτελεσματικά τις απαραίτητες πληροφορίες. Σε γενικές γραμμές, οι καταναλωτές είναι πιθανόν να έχουν κάποια

εμπειρία ή σχετικές πληροφορίες για τα προϊόντα που διαθέτουν ή σκοπεύουν να αγοράσουν (Chou et al. 2010).

Δεδομένου ότι η εποχή που διανύουμε είναι η εποχή της παγκοσμιοποίησης της οικονομίας και της ραγδαίας ανάπτυξης της τεχνολογίας, η βαθύτερη κατανόηση του πώς οι χρήστες θα μεταβαίνουν και θα περιηγούνται σε μία ιστοσελίδα είναι πολύ σημαντική για την ανάπτυξη των διαδικτυακών εφαρμογών, ιδίως όταν συμπεριλαμβάνει λειτουργίες όπως προτάσεις συμπληρωματικών προϊόντων, διαφημίσεις και εξατομικευμένες υπηρεσίες (Mobasher et al. 2002). Η ανάλυση των δεδομένων που προέρχονται από το κλικάρισμα του χρήστη, μπορεί να βοηθήσει στην κατανόηση της συμπεριφοράς πλοήγησής του, παρέχοντας λεπτομερείς πληροφορίες σχετικά με τα μοτίβα που δημιουργούνται από τους διάφορους χρήστες, δεδομένου ότι η πλοήγηση γίνεται σε ιστοσελίδα που φιλοξενείται στον ίδιο διακομιστή. Ωστόσο, το μέγεθος και η φύση των αρχείων αυτών μπορεί να καταστήσουν την ταξινόμηση και ανίχνευση μοτίβων σε δύσκολες και χρονοβόρες διαδικασίες.

Η ανάλυση των δεδομένων των χρηστών αποκαλύπτει σε πραγματικό χρόνο τις γεωγραφικές περιοχές στις οποίες η εταιρεία έχει ευκαιρίες για το ηλεκτρονικό επιχειρείν. Επίσης, παρέχει τρόπους στόχευσης των πελατών και κατανόησης των αναγκών τους, ώστε να προσαρμόσουν τις υπηρεσίες και στρατηγικές τους σε πραγματικό χρόνο. Επίσης, συνεισφέρει στην αποτελεσματικότητα ενός δικτυακού τόπου ως κανάλι του μάρκετινγκ, ποσοτικοποιώντας τις συμπεριφορές του χρήστη όσο βρίσκεται στην ιστοσελίδα (Purandare 2008).

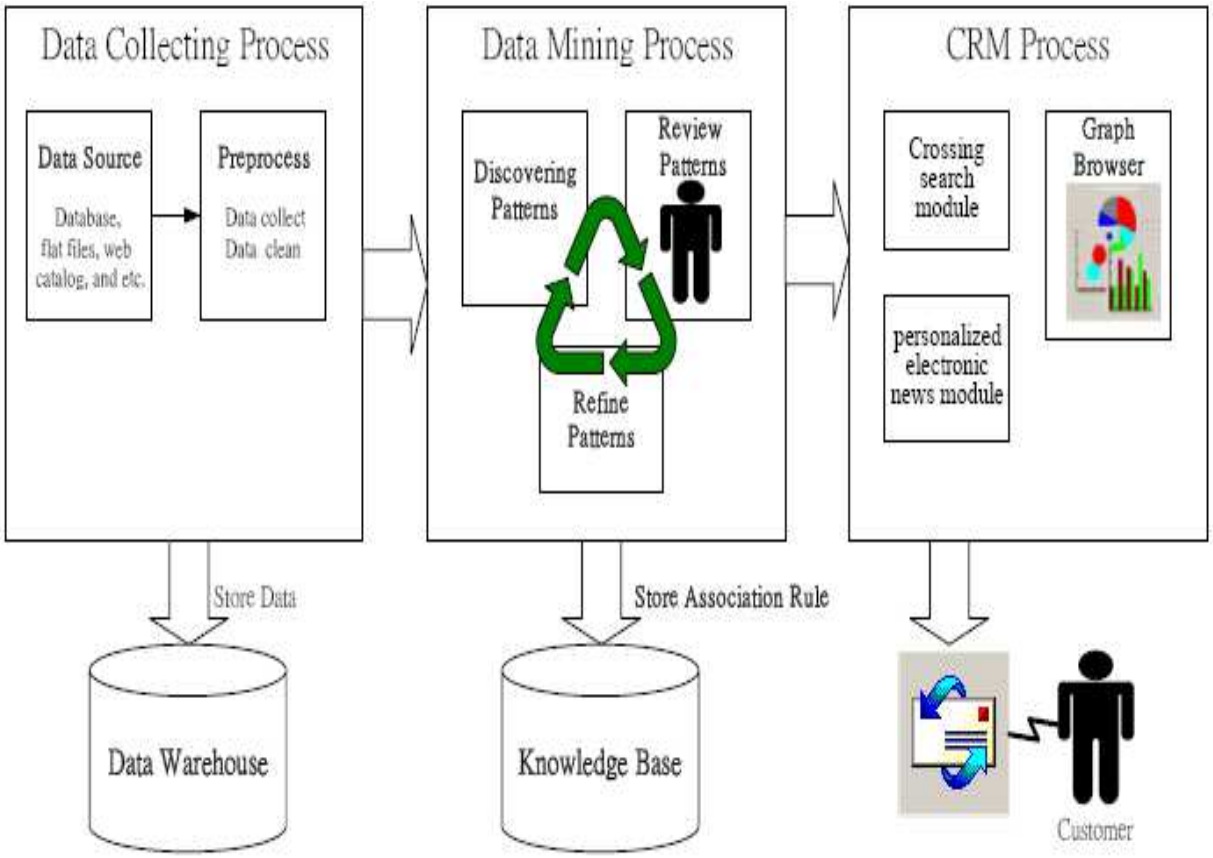
## 4.8 Το Web Mining στη διαχείριση πελατειακών σχέσεων

Δεδομένου ότι η online ενημέρωση έχει γίνει πολύ προσιτή και άφθονη, οι καταναλωτές ενημερώνονται συνεχώς, με αποτέλεσμα να γίνονται περισσότερο απαιτητικοί και εκλεπτυσμένοι. Για το λόγο αυτό, οι επιχειρήσεις πρέπει να διαφοροποιούν τα προϊόντα ή τις υπηρεσίες τους. Ένα μέσο επίτευξης της διαφοροποίησης αποτελεί η χρήση συστημάτων που αλληλεπιδρούν άμεσα με τους πελάτες, με συνέπεια και ακρίβεια, και απαιτεί την αποθήκευση των αρχείων των συναλλαγών και των συμπεριφορών των πελατών σε ένα online σύστημα, διαθέσιμο για ανάλυση σε έμπειρο και εξειδικευμένο προσωπικό (Rygielski et al. 2002).

Η εκτέλεση των αλγορίθμων εξόρυξης δεδομένων στα καταγεγραμμένα δεδομένα των χρηστών μπορεί να αποφέρει σημαντικές ανακαλύψεις για το είδος των προϊόντων που αγοράζονται συχνά μαζί. Αυτή η γνώση μπορεί στη συνέχεια να χρησιμοποιηθεί για τη δημιουργία συνδυασμών των προϊόντων και την υποβολή προτάσεων και συστάσεων προϊόντων σε κάποιον μελλοντικό πελάτη. Η διαχείριση πελατειακών σχέσεων χρησιμοποιεί τη γνώση που προήλθε μέσω της διαδικασίας της εξόρυξης δεδομένων και βοηθάει τους επισκέπτες τις ιστοσελίδας να λαμβάνουν καλύτερες αγοραστικές αποφάσεις σε μικρότερο χρονικό διάστημα. Βέβαια, η διαδικασία αυτή εξυπηρετεί και τους σκοπούς της επιχείρησης. Συγκεκριμένα, οι στόχοι του web usage mining για τους σκοπούς του μάρκετινγκ είναι οι εξής (Araya et al. 2004):

- Η απόκτηση νέων πελατών (acquire new customers).
- Η διατήρηση των υπαρχόντων πελατών (retain customers).
- Η πρόβλεψη αλλαγών στις προτιμήσεις των πελατών (predict churn of customers).
- Οι σταυροειδείς πωλήσεις (cross selling).
- Η πρόβλεψη μελλοντικών προτύπων-κανόνων επισκεψιμότητας των ιστοσελίδων (predict future visit patterns).

Οι Huang, Kwan και Li στο άρθρο τους (2003), προτείνουν ένα μοντέλο web mining το οποίο ενσωματώνει τη διαδικασία της διαχείρισης πελατειακών σχέσεων. Το μοντέλο αυτό απεικονίζεται στην *εικόνα 4.3*.




Εικόνα 4.3: Μοντέλο web mining στο CRM (Huang et al. 2003).

Όπως φαίνεται και στην *εικόνα 4.3*, το μοντέλο αποτελείται από τρία στάδια. Στο πρώτο στάδιο, γίνεται η συλλογή και ο καθαρισμός των δεδομένων, και στη συνέχεια η αποθήκευσή τους στις αποθήκες δεδομένων. Στο δεύτερο στάδιο, τα εργαλεία της εξόρυξης δεδομένων ενεργούν επαναληπτικές διαδικασίες, προκειμένου να εξαχθούν πρότυπα-μοτίβα, τα οποία θα οδηγήσουν στη δημιουργία ενός νέου συνόλου ερωτήσεων. Οι ερωτήσεις αυτές θα βελτιώσουν και θα καταστήσουν πιο λεπτομερή τη διαδικασία της έρευνας. Έπειτα, τα αποτελέσματα της εξόρυξης δεδομένων θα μεταφραστούν σύμφωνα με τους αποθηκευμένους κανόνες συσχέτισης. Τέλος, στο τρίτο στάδιο, τα εξαγόμενα πρότυπα-μοτίβα αποτελούν πλέον αξιόλογους δείκτες πρόβλεψης της καταναλωτικής συμπεριφοράς και οι διαδικασίες του CRM χρησιμοποιούν τα αποτελέσματα που παράγονται από τη διαδικασία εξόρυξης δεδομένων για την ενίσχυση της εστίασης στους πελάτες, αυξάνοντας έτσι τα ποσοστά ανταπόκρισης των πελατών και της αποτελεσματικότητας της στρατηγικής.

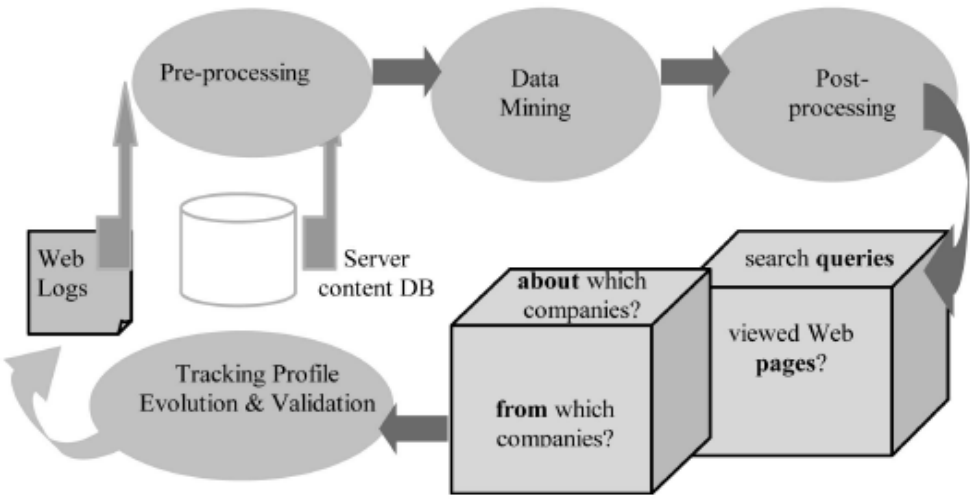
## 4.9 Ανακάλυψη του προφίλ πελατών μέσω του Web Usage Mining

Το web mining βοηθάει στην κατανόηση προβληματισμών, όπως η τρέχουσα και μελλοντική πιθανότητα διατήρησης του χρήστη ή τη σχέση μεταξύ συμπεριφοράς και αφοσίωσής του στην ιστοσελίδα. Τα μοντέλα που βασίζονται στη διαδικτυακή συμπεριφορά του πελάτη μπορούν να χρησιμοποιηθούν, όχι μόνο για τη βελτίωσης της τμηματοποίησης, αλλά και στην πρόβλεψη της μελλοντικής συμπεριφοράς του πελάτη, στοιχείο το οποίο είναι απαραίτητο στη σχεδίαση και τον προγραμματισμό του περιεχομένου της ιστοσελίδας. (Purandare 2008).

Η διαδικασία του web usage mining, προκειμένου να ανακαλυφθούν τα προφίλ των επισκεπτών μίας ιστοσελίδας, ξεκινάει με την ενοποίηση και την προεργασία των άμεσα εισερχόμενων δεδομένων στον διακομιστή (web server logs) με το περιεχόμενο της βάσης δεδομένων του διακομιστή, συνεχίζει με τον καθαρισμό των δεδομένων και στη συνέχεια ανακαλύπτονται τα μοτίβα και πρότυπα μέσω της συσταδοποίησης της εξόρυξης δεδομένων. Έπειτα, ακολουθεί δεύτερη επεξεργασία των συσταδοποιημένων, πλέον δεδομένων, ώστε να ανακαλυφθούν τα προφίλ των χρηστών. Τέλος, η διαδικασία καταλήγει στην παρακολούθηση της εξέλιξης των προφίλ, αφού νέα δεδομένα εισέρχονται καθημερινά και οι πληροφορίες που έχουμε απαρχαιώνονται συνεχώς (Nasraoui et al. 2008). Ο αυτόματος προσδιορισμός των προφίλ των χρηστών μέσω του web usage mining συνοψίζεται στα ακόλουθα βήματα:

1. Προ-επεξεργασία των web log αρχείων (pre-processing).
2. Συσταδοποίηση των συναλλαγών του χρήστη (clustering).
3. Σύνοψη των ομάδων/κατηγοριών των συναλλαγών στο προφίλ του χρήστη (summarize).
4. Εμπλουτισμός του προφίλ του χρήστη με πρόσθετα στοιχεία από επόμενες συναλλαγές του χρήστη αλλά και γνώσεις από εξωτερικές πηγές (enrich).
5. Παρακολούθηση των προφίλ (track current profiles).

Στην *εικόνα 4.4* απεικονίζεται τα πέντε αυτά βήματα της διαδικασία του web usage mining για την ανακάλυψη του προφίλ των χρηστών.




Εικόνα 4.4: Διαδικασία του Web Usage Mining για την ανακάλυψη του προφίλ των χρηστών (Nasraoui et al. 2008).

# ΚΕΦΑΛΑΙΟ 5: ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ

## 5.1 Εισαγωγή

Η εξόρυξη δεδομένων εφαρμόζεται σε ένα αρκετά ευρύ φάσμα βιομηχανιών και γίνεται όλο και πιο δημοφιλής λόγω της ουσιαστικής συμβολής της στην μείωση των δαπανών, καθώς επίσης και στην αύξηση του εισοδήματος. Επίσης, μπορεί να βοηθήσει τους οργανισμούς να κατανοήσουν τις ανάγκες της επιχείρησης, ώστε να είναι σε θέση να εξυπηρετήσουν τους πελάτες και να αυξήσουν την αποτελεσματικότητά τους σε μακροπρόθεσμη βάση. Στο κεφάλαιο αυτό αναλύονται οι εφαρμογές της εξόρυξης δεδομένων στην τουριστική βιομηχανία και στον κλάδο της υγείας. Επίσης, αναφέρονται περιληπτικά οι εφαρμογές της εξόρυξης δεδομένων στον τραπεζικό κλάδο, στο λιανικό εμπόριο και στις τηλεπικοινωνίες.

## 5.2 Εφαρμογή μεθόδων εξόρυξης δεδομένων στην τουριστική βιομηχανία

### 5.2.1 Εισαγωγή

Η τουριστική βιομηχανία αποτελεί έναν από τους μεγαλύτερους χρήστες της τεχνολογίας των πληροφοριών (IT), λόγω της φύσης των τουριστικών επιχειρήσεων (Magnini et al. 2003). Με την εξόρυξη δεδομένων και την ανακάλυψη γνώσης, χρησιμοποιώντας αποθηκευμένες και λεπτομερείς πληροφορίες για τους πελάτες, την αγορά και τις υπηρεσίες, οι τουριστικοί οργανισμοί και τουριστικοί πράκτορες θα είναι κατάλληλα εξοπλισμένοι ώστε να συμβαδίζουν και να αντιμετωπίζουν τις συνεχώς μεταβαλλόμενες ανάγκες της αγοράς (Pyo et al. 2002).

Η εξόρυξη δεδομένων και η ανάλυση διαδικτυακών δεδομένων (Wed data analysis) αποτελούν σημαντικά εργαλεία στην ηλεκτρονική διαχείριση πελατειακών σχέσεων στην τουριστική βιομηχανία. Η εξόρυξη δεδομένων αναφέρεται σε μία ετερογενή ομάδα στατιστικών τεχνικών και τεχνικών τεχνητής νοημοσύνης (AI), οι οποίες επιτρέπουν την εξαγωγή χρήσιμων πληροφοριών από μεγάλες βάσεις δεδομένων ή αποθήκες δεδομένων (Olmeda and Sheldon 2001). Το διαδίκτυο αποτελεί μία σύγχρονη βάση επικοινωνίας μεταξύ των ατόμων και, αν χρησιμοποιηθεί σωστά, μπορεί να αποτελέσει ένα ισχυρό εργαλείο για τους φορείς της τουριστικής βιομηχανίας.

Τα εξελιγμένα λογισμικά και η τεχνολογία των επικοινωνιών βοηθούν τους δυνητικούς και υπάρχοντες ταξιδιώτες να εντοπίσουν προορισμούς συνεπείς με τις προσδοκίες τους, τις προσωπικότητές τους και τις συνεχώς μεταβαλλόμενες ανάγκες τους (Pyo et al. 2002). Η σύγχρονη αποτελεσματική προώθηση τουριστικών προορισμών περιλαμβάνει ολοκληρωμένες γνώσεις, οι οποίες συμπεριλαμβάνουν χαρακτηριστικά του τουρισμού, εικόνες, συμπεριφορές και ιδιαιτερότητες των προτιμώμενων προορισμών. Στο μέλλον, οι τουριστικοί οργανισμοί θα αναγκαστούν να ακολουθήσουν μία περισσότερο επιθετική και στοχευμένη στρατηγική μάρκετινγκ (Buhalis 1999).

### 5.2.2 Η εξόρυξη δεδομένων στο διαδικτυακό τουριστικό μάρκετινγκ

Η επανάσταση στο διαδίκτυο αγγίζει κάθε πτυχή της ζωής μας, και ιδιαίτερα τη ζωή των ταξιδιωτών. Σύμφωνα με τους Watson, Zinkhan και Pitt (2000), το διαδίκτυο μπορεί να βοηθήσει σημαντικά τις στρατηγικές προκλήσεις που αντιμετωπίζουν οι τουριστικοί οργανισμοί. Συγκεκριμένα, τα πλεονεκτήματα από τη χρήση του διαδικτύου είναι τα εξής: μειωμένος κίνδυνος λόγω της μεταβαλλόμενης ζήτησης των τουριστικών υπηρεσιών (demand risk), ανάπτυξη νέων αγορών και προϊόντων με αποτέλεσμα τη διαφοροποίηση των τουριστικών προϊόντων και υπηρεσιών, μείωση του κινδύνου αναποτελεσματικότητας λόγω μειωμένου λειτουργικού κόστους και αυξημένου περιθωρίου κέρδους (inefficiency risk) και τέλος, αντιμετώπιση του κινδύνου καινοτομίας (innovation risk) που ενυπάρχει λόγω του ανταγωνισμού (Watson et al. 2000).

Οι τεχνικές της εξόρυξης δεδομένων μπορούν να εφαρμοστούν στην τουριστική βιομηχανία, και συγκεκριμένα στο διαδικτυακό μάρκετινγκ της τουριστικής βιομηχανίας, προσφέροντας πολλά πλεονεκτήματα, τα οποία είναι τα εξής (Olmeda and Sheldon 2001):

- Αυτόματος επανασχεδιασμός και ενημέρωση των ιστοσελίδων και των ηλεκτρονικών καταλόγων: Εάν ο χρήστης δεν βρει τις απαραίτητες πληροφορίες ή αν δαπανήσει πολύ χρόνο για να τις βρει, η ιστοσελίδα θα έχει αποτύχει στον στόχο για τον οποίο έχει σχεδιαστεί. Είναι, επίσης, σημαντικό να διασφαλιστεί ότι οι επόμενες ενημερώσεις των σελίδων θα ικανοποιούν τους περιορισμούς του σχεδιασμού που επιβάλλονται από τη συγκεκριμένη ιστοσελίδα.

- On-line δημοπρασίες: Η δημοπρασία υποτίθεται ότι είναι ο βέλτιστος μηχανισμός τιμολόγησης, κατά τον οποίο ανεξάρτητες προσφορές καταλήγουν σε ισορροπία των τιμών. Για το λόγο αυτό, οι χρήστες του διαδικτύου επηρεάζονται να αποδεχθούν ότι οι τιμές από τις δημοπρασίες είναι οι πιο κοντινές στην πραγματική αξία ενός αγαθού ή προϊόντος.

- Φιλτράρισμα των e-mails, εξυπηρέτηση πελατών μέσω e-mail και ανατροφοδότηση: Η χρήση του e-mail καθιστά γρήγορη και ανέξοδη την επικοινωνία με ένα μεγάλο αριθμό υφιστάμενων και δυνητικών πελατών.

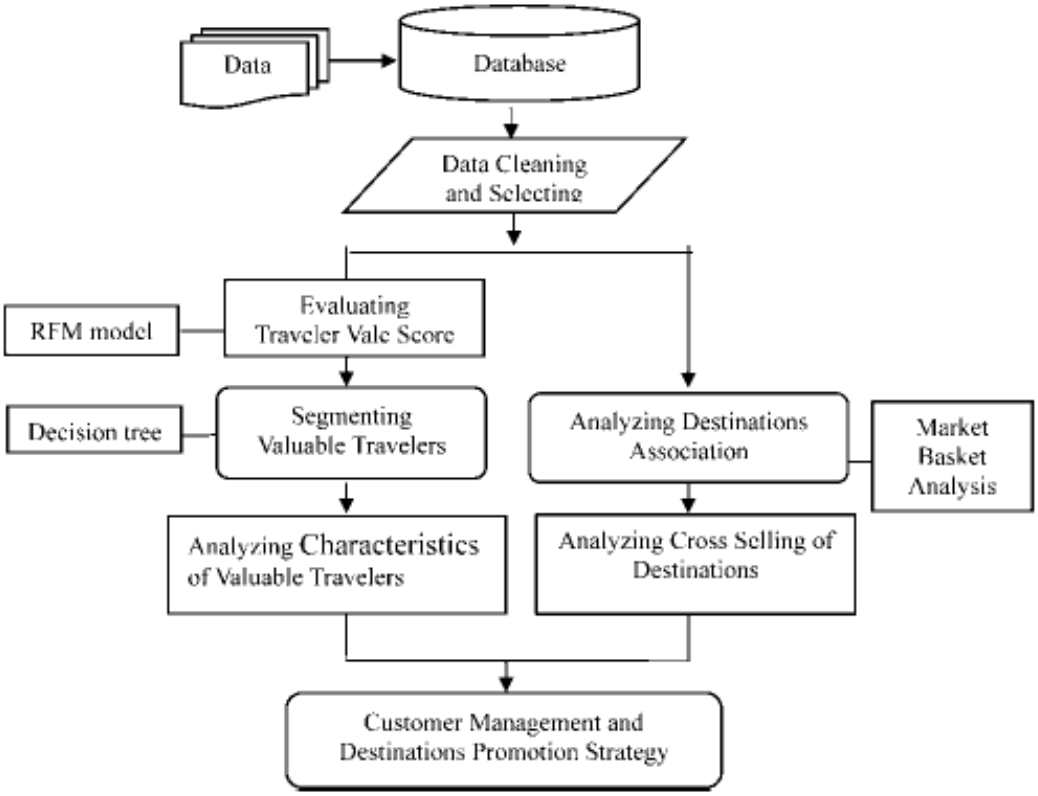
Επίσης, αποτελεί ένα εξαιρετικό μέσο για να παρέχει στους πελάτες πληροφορίες της τελευταίας στιγμής (π.χ. πιθανές καθυστερήσεις στις αεροπορικές εταιρείες). Τέλος, το e-mail είναι ένα αμφίδρομο κανάλι επικοινωνίας μεταξύ της εταιρείας και των πελατών (π.χ. έρευνα για το επίπεδο ικανοποίησης των πελατών από τις ταξιδιωτικές υπηρεσίες που εξέλαβαν).

### 5.2.3 Η διαδικασία της εξόρυξης δεδομένων στον τουριστικό οργανισμό

Σύμφωνα με τους Wong, Chen, Chung και Kao (2006), ο πιο σημαντικός στόχος για τους ταξιδιωτικούς οργανισμούς είναι ο εντοπισμός των πολύτιμων πελατών (valuable customers). Οι πολύτιμοι πελάτες είναι αυτοί που αποφέρουν τα μεγαλύτερα έσοδα στην εταιρεία. Κατά τον Wyner (1996), «το 80 τοις εκατό των εταιρικών κερδών δημιουργείται από μόλις το 20 τοις εκατό των πελατών».

Το δεύτερο πιο σημαντικό θέμα είναι να γνωρίζουν τον επόμενο πιθανό προορισμό τους. Χρησιμοποιώντας με ακρίβεια την μέθοδο των σταυροειδών πωλήσεων, ο ταξιδιωτικός οργανισμός θα μπορεί να παρέχει με ακρίβεια τις υπηρεσίες που χρειάζονται οι ταξιδιώτες, αλλά και θα υπάρχει η ευκαιρία προώθησης των πακέτων υπηρεσιών που οι πελάτες δεν έχουν αγοράσει ακόμα. Οι σταυροειδείς πωλήσεις μπορούν, επίσης, να χρησιμοποιηθούν στην αποτελεσματική διατήρηση των πελατών, ενισχύοντας την αφοσίωσή τους.

Εάν ο ταξιδιωτικός οργανισμός χρησιμοποιήσει τις τεχνικές εξόρυξης δεδομένων για να ανακαλύψει τα χαρακτηριστικά των πολύτιμων πελατών και τους προτιμώμενους επόμενους προορισμούς ταξιδίων, θα είναι πιο εύκολο να τμηματοποιήσει τους ταξιδιώτες για τους σκοπούς των σταυροειδών πωλήσεων. Η διαδικασία εξόρυξης δεδομένων που σχεδιάστηκε από τους Wong, Chen, Chung και Kao (2006) για την επίτευξη των δύο αυτών στόχων απεικονίζεται στην *εικόνα 5.1*.




Εικόνα 5.1: Διαδικασία εξόρυξης δεδομένων στον ταξιδιωτικό οργανισμό (Wong et al. 2006).

Όπως φαίνεται και στην *εικόνα 5.1*, οι πληροφορίες των πελατών αποθηκεύονται στην αποθήκη δεδομένων και στη συνέχεια τα δεδομένα «καθαρίζονται» (data cleaning). Στο επόμενο βήμα χρησιμοποιείται το RFM μοντέλο ώστε να αξιολογηθούν οι πολύτιμοι ταξιδιώτες, εφαρμόζοντας το δέντρο αποφάσεων για να ανακαλυφθούν τα χαρακτηριστικά των πολύτιμων πελατών. Στο τρίτο βήμα, μέσω της ανάλυσης του καλαθιού αγοράς, συλλέγονται οι προορισμοί που έχουν επισκεφθεί οι πελάτες, και υποβάλλονται σε ανάλυση (association analysis) ώστε να προβλεφθούν οι μελλοντικοί πιθανοί προορισμοί. Τέλος, επιλέγεται η βέλτιστη στρατηγική προώθησης των τουριστικών προορισμών για τους σκοπούς των σταυροειδών πωλήσεων. Τα αποτελέσματα από το δέντρο αποφάσεων και την ανάλυση του καλαθιού αγοράς επιτρέπουν στους ταξιδιωτικούς οργανισμούς να επικεντρώνονται στον ταξιδιώτη-στόχο και να προτείνουν μία στρατηγική διαχείρισης των πολύτιμων ταξιδιωτών.

## 5.3 Εφαρμογή μεθόδων εξόρυξης δεδομένων στη βιομηχανία της υγείας

### 5.3.1 Εισαγωγή

Η εξόρυξη δεδομένων χρησιμοποιείται εντατικά και εκτεταμένα στην υγειονομική περίθαλψη και μάλιστα γίνεται ολοένα και περισσότερο δημοφιλής, αν όχι απαραίτητη (Koh and Tan 2005). Οι εφαρμογές της εξόρυξης δεδομένων ωφελεί όλα τα εμπλεκόμενα μέλη στην βιομηχανία της υγείας. Παραδείγματα είναι τα εξής: οι ασφαλιστές υγείας μπορούν να ανιχνεύσουν την απάτη και κατάχρηση, οι οργανισμοί υγειονομικής περίθαλψης λαμβάνουν ευκολότερα αποφάσεις διαχείρισης πελατειακών σχέσεων, οι γιατροί προσδιορίζουν αποτελεσματικές θεραπείες, και οι ασθενείς λαμβάνουν καλύτερες και οικονομικά προσιτές υπηρεσίες υγείας. Οι τεράστιες ποσότητες δεδομένων που προκύπτουν από τις συναλλαγές στην υγειονομική περίθαλψη αποδεικνύονται υπερβολικά πολύπλοκες και ογκώδεις για να επεξεργαστούν και να αναλυθούν από τις παραδοσιακές μεθόδους (Viswanathan 2007) Η εξόρυξη δεδομένων παρέχει την απαραίτητη μεθοδολογία και την τεχνολογία για να μετατρέψει τους τεράστιους αυτούς όγκους δεδομένων σε χρήσιμες πληροφορίες για τη λήψη αποφάσεων.

Οι κλάδοι της υγειονομικής περίθαλψης επιλέγουν πλέον τη διαχείριση πελατειακών σχέσεων για την κατασκευή μίας γέφυρας εμπιστοσύνης μεταξύ των νοσοκομείων και των πελατών, έτσι ώστε οι πελάτες να μην αισθάνονται αποξενωμένοι στο νοσοκομειακό περιβάλλον. Το CRM επιτρέπει στην βιομηχανία της υγείας να λαμβάνει σημαντικές πληροφορίες για τους πελάτες και να τις χρησιμοποιούν όσο το δυνατόν αποτελεσματικότερα (Yina 2010).

Στην *εικόνα 5.2* απεικονίζονται τα επίπεδα εφαρμογών της εξόρυξης δεδομένων στην βιομηχανία της υγείας. Όπως φαίνεται και στην εικόνα, υπάρχουν τρία επίπεδα εφαρμογών, το χαμηλό επίπεδο (low level), το νοσοκομειακό επίπεδο (hospital-based) και το παγκόσμιο επίπεδο (global level). Η εφαρμογή της εξόρυξης δεδομένων στη διαχείριση πελατειακών σχέσεων αναφέρεται στο νοσοκομειακό επίπεδο.

| | |
|---|---|
| Low Level | sensor data mining for emergency detection |
| | sensor data fusion for decision support |
| Hospital-Based | periodic trend analysis & wellness recommendations |
| | analysis of patient groups for drug effects, CRM and indirect factors |
| | recovery analysis, drug interaction and other studies like allergy cause and detection |
| Global Level | clustering and classification analysis for epidemics and other implicit patterns |
| | GRID based mining of biomedical data like gene expressions and MRI images, on a National and International scale. |

Εικόνα 5.2: Επίπεδα εφαρμογής της εξόρυξης δεδομένων στη βιομηχανία της υγείας (Viswanathan, 2007).

Σύμφωνα με τις εκτιμήσεις του Yina (2010), οι υπηρεσίες υγειονομικής περίθαλψης αντιπροσωπεύουν περίπου το 10% του ΑΕΠ των δυτικών εθνών. Αυτό το ποσοστό περιλαμβάνει τις πληρωμές προς τα νοσοκομεία, τους γιατρούς, τα γηροκομεία, τα διαγνωστικά εργαστήρια, τις φαρμακευτικές υπηρεσίες και τους κατασκευαστές ιατρικών συσκευών. Παρακάτω γίνεται ανάλυση των εφαρμογών της εξόρυξης δεδομένων στο νοσοκομειακό επίπεδο.

### 5.3.2 Εφαρμογή της εξόρυξης δεδομένων στη διαχείριση πελατειακών σχέσεων της νοσοκομειακής μονάδας

Η διαχείριση πελατειακών σχέσεων είναι ένας πολύ σημαντικός τομέας στην υγειονομική περίθαλψη. Οι αλληλεπιδράσεις με τους πελάτες μπορούν να προκύψουν μέσα από τα τηλεφωνικά κέντρα, τα γραφεία των γιατρών, τα τμήματα τιμολόγησης και τα κέντρα εσωτερικής παραμονής (Koh and Tan 2005). Όπως και στην περίπτωση των εμπορικών εταιρειών, οι εφαρμογές της εξόρυξης δεδομένων στον κλάδο της υγείας αναπτύσσονται με σκοπό να καθοριστούν οι προτιμήσεις, η συχνότητα χρήσης, καθώς και οι τρέχουσες και μελλοντικές ανάγκες των ατόμων, ώστε να βελτιωθεί το επίπεδο ικανοποίησης (Biafore 1999). Οι εφαρμογές αυτές, μπορούν επίσης να

χρησιμοποιηθούν στην πρόβλεψη άλλων προϊόντων ή υπηρεσιών, των οποίων είναι πιθανό να κάνει χρήση ο ασθενής.

Η βάση για ένα CRM πρόγραμμα υγειονομικής περίθαλψης είναι η ίδια όπως και για κάθε άλλη επιχείρηση. Απαιτεί ένα ολοκληρωμένο συνδυασμό δεδομένων και προγραμμάτων εφαρμογών για την υποστήριξη της ανάλυσης, την αναγνώριση ευκαιριών, την εξόρυξη γνώσης και τη διαχείριση των επικοινωνιών (Hallick 2001). Ωστόσο, οι διαφορές στον τομέα της υγείας είναι οι συγκεκριμένες πληροφορίες που απαιτούνται για την κατασκευή της βάσης δεδομένων, όπως για παράδειγμα οι τεχνικές ορολογίες και τα διαδικαστικά και διαγνωστικά ακρωνύμια στα δεδομένα.

Μία ολοκληρωμένη αποθήκη δεδομένων μπορεί να παρέχει μία ολοκληρωμένη εικόνα των ασθενών, η οποία περιλαμβάνει τις δραστηριότητές τους, τα χαρακτηριστικά και τις συμπεριφορές σε επίπεδο νοικοκυριού, αλλά και τα ακόλουθα είδη πληροφοριών (Hallick 2001):

- ✓ εσωτερικές πληροφορίες, οι οποίες συλλέγονται από τα λογιστικά και ιατρικά αρχεία του οργανισμού,
- ✓ εξωτερικές πληροφορίες, οι οποίες συγκεντρώνονται από εξωτερικές πηγές δεδομένων, και τέλος
- ✓ υπολογιστικές πληροφορίες για την ανάλυση των δεδομένων, την πρόβλεψη της αγοράς, τους κινδύνους για την υγεία, και άλλες προβλέψεις.

Τα εργαλεία ανάλυσης μπορούν να αναζητήσουν δεδομένα στην CRM βάση δεδομένων για να απαντήσουν σε ερωτήσεις που στο παρελθόν η απάντησή τους ήταν άγνωστη, ή δε μπορούσαν να απαντηθούν γρήγορα. Τέτοιου είδους ερωτήσεις θα μπορούσαν να είναι οι ακόλουθες: Ποιοι πελάτες στη βάση δεδομένων μας έχουν καλέσει στο τηλεφωνικό κέντρο, πόσοι επισκέφθηκαν την ιστοσελίδα μας, πόσοι έχουν κάνει χρήση των παρεχόμενων υπηρεσιών, και άλλα.

Ένα CRM σύστημα χρειάζεται χρόνο και προετοιμασία για να ολοκληρωθεί, όπως επίσης απαραίτητη είναι και η καθημερινή συσσώρευση δεδομένων, τόσο από το νοσοκομειακό επίπεδο όσο και από τα εξωνοσοκομειακά τμήματα. Απαραίτητη κρίνεται και η ενοποίηση του CRM με άλλα βοηθητικά τεχνικά συστήματα. Ένα αποτελεσματικό CRM σύστημα θεωρείται ολοκληρωμένο αφού έχουν ολοκληρωθεί τα παρακάτω βήματα (Yina 2010):

1. Ενσωμάτωση του CRM συστήματος με το σύστημα HIS (Hospital Information System): Το CRM σύστημα είναι η επέκταση και αλληλοσυμπλήρωση του Νοσοκομειακού Πληροφοριακού Συστήματος (HIS). Δεδομένου ότι τα ιατρικά δεδομένα αποτελούν τη βάση για την διαχείριση πελατειακών σχέσεων στα νοσοκομεία, το CRM χρησιμοποιεί αναμφίβολα ως κύρια πηγή δεδομένων το HIS και για το λόγο αυτό θα πρέπει να εξασφαλισθεί η σταθερότητά του. Τα δεδομένα στη βάση δεδομένων του HIS μπορούν να χρησιμοποιηθούν ανά πάσα στιγμή για την ανάλυση και την εξόρυξη δεδομένων. Με βάση την ανάλυση των δεδομένων του HIS, το CRM μπορεί να εντοπίσει το κατάλληλο σύστημα διαχείρισης και το μοντέλο μάρκετινγκ που ταιριάζει στο νοσοκομείο, όπως επίσης και τις βασικές ομάδες πελατών και τις ειδικές υπηρεσίες που απαιτούν.

2. Ενσωμάτωση του CRM συστήματος με τη διαδικτυακή νοσοκομειακή πλατφόρμα (web platform): Οι ασθενείς συνήθως επισκέπτονται το νοσοκομείο όταν χρειάζονται εξειδικευμένες κλινικές υπηρεσίες, το οποίο συνεπάγεται ότι έχει κανονιστεί κάποιο ραντεβού μία με δύο μέρες νωρίτερα. Η ενσωμάτωση με τη διαδικτυακή νοσοκομειακή πλατφόρμα μπορεί να μειώσει τη συχνότητα επίσκεψης των ασθενών στο νοσοκομείο, με τη δημιουργία μιας on-line κράτησης και το πρόγραμμα αποτελεσμάτων (check-up) για τον εξορθολογισμό της κλινικής διαδικασίας.

3. Ενσωμάτωση του CRM με το τηλεφωνικό κέντρο: Οι παραδοσιακές τηλεφωνικές υπηρεσίες των νοσοκομείων παρέχουν μόνο στατικές πληροφορίες σχετικά με τα κλινικά τμήματα, του γιατρούς, τις ιατρικές αμοιβές κλπ. Η ενσωμάτωση του CRM με το τηλεφωνικό κέντρο διευκολύνει τους χειριστές του τηλεφωνικού κέντρου ώστε, όταν λαμβάνουν κάποια κλήση να γνωρίζουν τις απαραίτητες πληροφορίες για τους ασθενείς, όπως την ταυτότητά τους, το ιατρικό ιστορικό τους, και άλλα. Με αυτόν τον τρόπο οι χειριστές είναι σε θέση να δίνουν ακριβείς και έγκαιρες απαντήσεις στους ασθενείς και να κανονίζουν τα ραντεβού με τους κατάλληλους γιατρούς και ειδικούς.

4. Ενσωμάτωση του CRM με το μηχανισμό σύντομων μηνυμάτων (short-message gateway): Ως ένα αποτελεσματικό συμπληρωματικό δίκτυο επικοινωνίας, ο μηχανισμός σύντομων μηνυμάτων παίζει σημαντικό ρόλο στη μετάβαση

πληροφοριών από το CRM στα νοσοκομεία. Στην πράξη, υπηρεσίες όπως η υπενθύμιση μέσω μηνυμάτων, επιβεβαίωση των ραντεβού και η φροντίδα των ασθενών μπορεί να προσφερθεί στους πελάτες με την αποστολή τέτοιου είδους μηνυμάτων.

5. Δημιουργία on-line πλατφόρμας αλληλεπίδρασης μεταξύ του νοσοκομείου και των ασθενών: Η διαδικτυακή πλατφόρμα του νοσοκομείου μπορεί να χρησιμοποιηθεί ώστε να να συλλέγει τις συμβουλές και καταγγελίες από τους ασθενείς, να παρέχει on-line ιατρικές διαβουλεύσεις και να προσφέρει υπηρεσίες κράτησης με γιατρούς και ειδικούς για τους εξωνοσοκομειακούς ασθενείς. Στη συνέχεια, οι πληροφορίες που συλλέγονται στην διαδικτυακή πλατφόρμα θα εισέλθουν στο σύστημα CRM, όπου το αρμόδιο προσωπικό ειδοποιεί τα σχετικά τμήματα για να προβούν στην παρακολούθηση των πληροφοριών αυτών.

6. Δημιουργία μηχανισμού ανταπόκρισης στον πελάτη και δημιουργία βάσης δεδομένων: Η επεξεργασία των παραπόνων των ασθενών αποτελεί ένα σημαντικό μέρος της ρουτίνας των νοσοκομείων. Ο μηχανισμός ενσωμάτωσης του CRM με το σύστημα που υποστηρίζει το διαδίκτυο, το τηλέφωνο και το ταχυδρομείο παρέχει μία πολυδιάστατη πλατφόρμα διαχείρισης παραπόνων, όπου οι συμβουλές και καταγγελίες τηρούνται σε αρχεία, με άμεσο αποτέλεσμα την έγκαιρη ανταπόκριση ώστε να ενισχύεται η ικανοποίηση των ασθενών. Με αυτόν τον τρόπο επιτυγχάνεται η τυποποίηση της διαχείρισης του νοσοκομείου.

### 5.3.3 Άλλες κατηγορίες εφαρμογής μεθόδων εξόρυξης δεδομένων στο νοσοκομειακό επίπεδο

Σύμφωνα με τους Koh και Tan (2005), άλλες κατηγορίες εφαρμογής της εξόρυξης δεδομένων στο νοσοκομειακό επίπεδο, πέραν της διαχείρισης πελατειακών σχέσεων, είναι οι ακόλουθες:

- Αποτελεσματικότητα της θεραπείας: Συγκρίνοντας και αντιπαραθέτοντας τις αιτίες, τα συμπτώματα και τις θεραπευτικές αγωγές, η εξόρυξη δεδομένων μπορεί να εξάγει μία ανάλυση με τα πιο αποτελεσματικά σχέδια δράσης. Άλλες

εφαρμογές της εξόρυξης δεδομένων που σχετίζονται με τις θεραπείες αποτελούν η ομαδοποίηση των διάφορων παρενεργειών τις θεραπείας, η συγκέντρωση των συμπτωμάτων για τη διάγνωση της ασθένειας, καθορισμός προληπτικών μέτρων ενάντια κάποιων ασθενειών και άλλα.

- Διαχείριση της υγειονομικής περίθαλψης: Για την ενίσχυση της διαχείρισης της υγειονομικής περίθαλψης, οι εφαρμογές εξόρυξης δεδομένων μπορούν να αναπτυχθούν ώστε να εντοπίζουν και να παρακολουθούν τις χρόνιες παθολογικές καταστάσεις και τους ασθενείς υψηλού κινδύνου, να σχεδιάζουν τις κατάλληλες επεμβάσεις και να μειώνουν τον αριθμό των εισαγωγών των ασθενών στο νοσοκομείο.

- Απάτη και κατάχρηση: Οι εφαρμογές της εξόρυξης δεδομένων που επιχειρούν να ανιχνεύσουν την απάτη και την κατάχρηση, δημιουργούν κανόνες και στη συνέχεια εντοπίζουν ασυνήθιστες ή μη φυσιολογικής μορφής απαιτήσεις από γιατρούς, κλινικές, εργαστήρια ή άλλους συνεργάτες. Είναι σε θέση, επίσης, να επισημάνουν ακατάλληλες συνταγές από ιατρούς ή ψευδείς ιατρικές ασφαλίσεις ασθενών.

## 5.4 Εφαρμογή της εξόρυξης δεδομένων σε άλλους κλάδους

### 5.4.1 Εφαρμογή της εξόρυξης δεδομένων στον τραπεζικό τομέα

Τα εργαλεία της εξόρυξης δεδομένων εφαρμόζονται ευρέως στον τραπεζικό κλάδο. Συγκεκριμένα, εφαρμόζονται στην τμηματοποίηση πελατών, στη βαθμολόγηση της πιστοληπτικής ικανότητας, την έγκριση των τραπεζικών δανείων, την πρόβλεψη αθέτησης πληρωμής, την ανίχνευση παράνομων συναλλαγών, την διαχείριση των διαθεσίμων και τη βελτιστοποίηση των χαρτοφυλακίων μετοχών (Thanuja et al. 2011). Πέρα από την κατανόηση της αξία του πελάτη, η τράπεζα επωφελείται από τη δημιουργία καλύτερων πελατειακών σχέσεων, αυξάνοντας παράλληλα την αφοσίωση των πελατών και, κατά συνέπεια, των εσόδων (Hsieh 2004).

Ωστόσο, η ανάλυση των βάσεων δεδομένων των τραπεζών πολλές φορές αποτελεί ένα δύσκολο εγχείρημα, καθώς τα τραπεζικά δεδομένα είναι πολυδιάστατα. Αποτελούνται

από αρχεία καθημερινών συναλλαγών και αρχεία λογαριασμών που ενημερώνονται μηνιαίως. Ως εκ τούτου, ακόμη και τα εξαιρετικά ακριβή μοντέλα εξόρυξης δεδομένων εμφανίζουν συχνά εσφαλμένους κανόνες τμηματοποίησης πελατών (Hsieh 2004).

### 5.4.2 Εφαρμογή της εξόρυξης δεδομένων στο λιανικό εμπόριο

Οι επιχειρήσεις του κλάδου του λιανικού εμπορίου δύνανται, μέσω των εργαλείων της εξόρυξης δεδομένων, να αποκτήσουν ανταγωνιστικό πλεονέκτημα, έναντι των άλλων επιχειρήσεων. Συγκεκριμένα, η εξόρυξη δεδομένων μπορεί να χρησιμοποιηθεί στην πρόβλεψη των πωλήσεων της επιχείρησης, τις εποχιακές διακυμάνσεις, ή ακόμα και στην πρόβλεψη των γενικών τάσεων της αγοράς. Επίσης, είναι δυνατή η παροχή πληροφοριών σχετικά με τις αγοραστικές συνήθειες και τις προτιμήσεις των πελατών, τον χρόνο παράδοσης των προϊόντων από τους προμηθευτές, και άλλες σχετικές πληροφορίες, οι οποίες συμβάλλουν στη λήψη στρατηγικών αποφάσεων. Οι πιο διαδεδομένοι τομείς εφαρμογής της εξόρυξης δεδομένων στον κλάδο του λιανικού εμπορίου είναι η διαχείριση του κινδύνου, η ανίχνευση της απάτης, η απόκτηση πελατών και η διατήρησή τους (Thanuja et al. 2011).

### 5.4.3 Εφαρμογή της εξόρυξης δεδομένων στον κλάδο τηλεπικοινωνιών

Μετά από χρόνια ύπαρξης μονοπωλίου στην αγορά, ο κλάδος των τηλεπικοινωνιών είναι πλέον ιδιαίτερα ανταγωνιστικός και οι πελάτες είναι σε θέση να στραφούν σε άλλον πάροχο πολύ εύκολα. Για το λόγο αυτό, οι εταιρείες τηλεπικοινωνιών διερευνούν τρόπους, μέσω των εργαλείων εξόρυξης δεδομένων, για να αποκτήσουν ανταγωνιστικό πλεονέκτημα. Μέσω της κατανόησης των χαρακτηριστικών και των συμπεριφορών των πελατών, οι εταιρείες είναι σε θέση να αποκτήσουν πιστούς πελάτες και να βελτιώσουν την κερδοφορία τους (Thanuja et al. 2011).

Ο κλάδος των τηλεπικοινωνιών είναι από τους πρώτους που υιοθέτησαν την τεχνολογία της εξόρυξης δεδομένων, κυρίως λόγω της μεγάλης ποσότητας και ποιότητας των δεδομένων που συλλέγει. Τα δεδομένα αυτά περιλαμβάνουν λεπτομερείς πληροφορίες κλήσεων, πληροφορίες του δικτύου και πληροφορίες για τους πελάτες. Η ποσότητα των δεδομένων είναι τόσο μεγάλη, ώστε οι αναλύσεις με απλά εργαλεία αναλύσεων

καθίσταται δύσκολη. Η εξόρυξη δεδομένων στον κλάδο των τηλεπικοινωνιών εφαρμόζεται σε τρεις κύριους τομείς: την ανίχνευση της απάτης, τη δημιουργία προφίλ πελατών και την απομόνωση βλαβών του δικτύου (Weiss 2005).

# ΚΕΦΑΛΑΙΟ 6: ΗΘΙΚΑ ΖΗΤΗΜΑΤΑ ΣΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

## 6.1 Εισαγωγή

Οι σύγχρονες εξελίξεις στην τεχνολογία έχουν επιτρέψει τη συλλογή και την επεξεργασία τεράστιων ποσοτήτων προσωπικών δεδομένων. Οι πληροφορίες αυτές είναι αναμφίβολα ιδιαίτερα χρήσιμες σε πολλούς τομείς. Ωστόσο, η εξόρυξη δεδομένων αποτελεί απειλή για ορισμένες σημαντικές ηθικές αξίες, όπως η προστασία της ιδιωτικής ζωής. Οι υφιστάμενοι νόμοι περί προστασίας της ιδιωτικότητας και των πολιτών έχουν απαρχαιωθεί από τις εξελίξεις τις τεχνολογίας και δεν προσφέρουν πλέον επαρκή προστασία. Για το λόγο αυτό, υπάρχει μία αυξανόμενη ανησυχία των πολιτών για την προστασία της προσωπικής ζωής των ατόμων. Συγκεκριμένα, σύμφωνα με έρευνα που έγινε για τη διερεύνηση της στάσης των καταναλωτών απέναντι στη χρήση δευτερογενών πληροφοριών (Culnan 1993), το 78% των ερωτηθέντων πιστεύει πως η τεχνολογία των υπολογιστών αποτελεί σημαντική απειλή της ιδιωτικής ζωής και ότι η χρήση των ηλεκτρονικών υπολογιστών θα πρέπει να περιοριστεί αισθητά στο μέλλον, προκειμένου να διαφυλαχθεί το ιδιωτικό απόρρητο.

Στο παρόν κεφάλαιο, αναλύονται οι έννοιες της ηθικής και της προστασίας των προσωπικών δεδομένων και αναφέρονται κάποιες κατευθυντήριες γραμμές του ΟΟΣΑ για την διαφύλαξη του απορρήτου. Στη συνέχεια, γίνεται σύγκριση των παραδοσιακών μεθόδων ανάκτησης πληροφοριών με τις τεχνικές της εξόρυξης δεδομένων και αναφέρονται οι νέες απειλές που προέρχονται από την εφαρμογή των εργαλείων της εξόρυξης στα προσωπικά δεδομένα. Τέλος, παρατίθενται ορισμένες τεχνικές για την ενίσχυση της προστασίας των δεδομένων.

## 6.2 Έννοιες – Ορισμοί

Η ανάπτυξη της εξόρυξης δεδομένων παρουσιάζει σημαντικά ηθικά και κοινωνικά ζητήματα που πρέπει να αντιμετωπιστούν, όταν οι τεχνολογικές εξελίξεις έχουν τύχει ευρείας αποδοχής. Η ηθική αποτελείται από τις αρχές που χρησιμοποιούμε για να διεκδικήσουμε την κοινωνική ευθύνη και να διαιωνίσουμε ασφαλείς και δίκαιες κοινωνίες (Wahlstrom et al. 2002). Η εγκυρότητα των κανόνων που παράγονται από τη διαδικασία της εξόρυξης δημιουργεί ένα ηθικό ζήτημα, όταν τα αποτελέσματα χρησιμοποιούνται στις διαδικασίες λήψης αποφάσεων που επηρεάζουν τους ανθρώπους, ενώ η εξόρυξη των δεδομένων των πελατών μπορεί μη ηθελημένα να θέσει σε κίνδυνο την προστασία της ιδιωτικής ζωής των πελατών (Roddick and Fule 2003).

Η σύγχρονη ανάγκη για μία πολύ μεγάλη βάση δεδομένων και τα συναφή εργαλεία αυτόματης ανάλυσης δεδομένων συνάγουν επιπρόσθετα ζητήματα ηθικής που πρέπει να εξεταστούν. Συγκεκριμένα, τα άτομα παράγουν ηλεκτρονικά δεδομένα με πολλούς τρόπους, όπως αγορές με πιστωτικές κάρτες, συνδρομές, ταχυδρομικές παραγγελίες, online αγορές, αρχεία τραπεζών, κ.λπ.. Όταν αυτά τα δεδομένα συγχωνεύονται και αναλύονται μπορούν να αποκαλύψουν τις προτιμήσεις των ατόμων, διάφορες συσχετίσεις, πληροφορίες πίστωσης, το ύψος του εισοδήματος, ακόμα και τις πολιτικές τους πεποιθήσεις (Wahlstrom et al. 2002).

## 6.3 Προστασία των προσωπικών δεδομένων

Η προστασία προσωπικών δεδομένων, η οποία συχνά συνδέεται και περιγράφεται με τους όρους ελευθερία, αυτονομία και απόρρητο, είναι μία έννοια που δεν είναι εύκολο να ορισθεί επακριβώς. Η προστασία προσωπικών δεδομένων συνήθως θεωρείται ως το δικαίωμα του ατόμου για τον έλεγχο των πληροφοριών που αφορούν τον εαυτό του (Brankovic and Estivill-Castro 1999). Ένα άτομο μπορεί να διατηρήσει την ζωή του προσωπική με τον περιορισμό της πρόσβασης σε αυτή από τρίτους, και αυτό επιτυγχάνεται καλύτερα με τον περιορισμό της διαθεσιμότητας των προσωπικών του πληροφοριών (Wahlstrom et al. 2002). Η προστασία της ιδιωτικής ζωής μπορεί να παραβιαστεί όταν οι πληροφορίες σχετικά με ένα άτομο λαμβάνονται, χρησιμοποιούνται ή διαδίδονται από τρίτους, ιδιαίτερα αν αυτό συμβαίνει χωρίς τη γνώση ή τη συγκατάθεσή του.

Σύμφωνα με τον Laudon (1996), τουλάχιστον 400 εκατομμύρια αρχεία πίστωσης, 700 εκατομμύρια αρχεία φαρμακευτικών αγωγών, 100 εκατομμύρια ιατρικά αρχεία και 600 εκατομμύρια προσωπικών αρχείων πωλούνται κάθε χρόνο στις ΗΠΑ. Ανάμεσα στα αρχεία που πωλούνται είναι τα τραπεζικά ισοζύγια, ιστορικά ενοικίασης, λιανικές αγορές, ποινικά μητρώα, τηλεφωνικοί αριθμοί και πρόσφατες τηλεφωνικές κλήσεις. Όλες αυτές οι πληροφορίες συμβάλλουν στην ανάπτυξη «εικόνων» των δεδομένων των ατόμων, και πωλούνται απευθείας σε επιχειρήσεις, ιδιώτες, ερευνητές και κρατικούς οργανισμούς.

Υπάρχουν τουλάχιστον δύο σημαντικοί τρόποι με τους οποίους η εισαγωγή μιας νέας τεχνολογίας μπορεί να εγείρει ανησυχίες σχετικά με την προστασία της ιδιωτικής ζωής (Tavani 1999):

1. χρησιμοποιείται τεχνολογία για τη συλλογή πληροφοριών ενός ατόμου ή μίας ομάδας ατόμων, χωρίς όμως να έχει την επίγνωση το άτομο ή η ομάδα ατόμων, και
2. τα άτομα γνωρίζουν ότι συλλέγονται πληροφορίες για τους ίδιους μέσω μίας συγκεκριμένης τεχνολογίας, αλλά δεν μπορούν να εκφέρουν κανένα λόγο για το πώς θα χρησιμοποιηθούν αυτές οι πληροφορίες, εάν δηλαδή θα αποκαλυφθούν, θα πωληθούν, θα ανταλλαχθούν, κ.λπ..

Τα εργαλεία εξόρυξης δεδομένων, από τη φύση τους, αποκαλύπτουν για τα άτομα πληροφορίες, οι οποίες θα ήταν εξαιρετικά δύσκολο να προβλεφθούν από τους χρήστες των προγραμμάτων, αλλά και να δώσουν τη συγκατάθεσή τους για αυτές τις πληροφορίες οι ίδιοι οι πελάτες.

Η προστασία της ιδιωτικής ζωής είναι κατανοητή με διαφορετικό τρόπο στις ΗΠΑ, την Ευρώπη και αλλού και η νομοθετική κατηγορία της προστασίας των δεδομένων δεν είναι αναγνωρισμένη το ίδιο σε πολλές χώρες. Αν λάβουμε υπόψιν μας τις τεχνολογικές δυνατότητες και τα εμπορικά συμφέροντα, καθίσταται αναγκαία η ύπαρξη ενός ισχυρού νομικού καθεστώτος (Birnhack 2008). Το 1980 ο Οργανισμός για Οικονομική Συνεργασία και Ανάπτυξη (ΟΟΣΑ – OECD) ενέκρινε τις κατευθυντήριες γραμμές για την προστασία των προσωπικών δεδομένων. Αυτές οι οδηγίες αναγνωρίζουν και ενθαρρύνουν το προνόμιο του ατόμου να ελέγχει τα προσωπικά του στοιχεία. Δύο από αυτές τις κατευθυντήριες γραμμές ορίζουν ότι ο λόγος για τη συλλογή προσωπικών δεδομένων θα πρέπει να καταστεί γνωστός και σαφής στο άτομο πριν από τη συλλογή

τους και ότι τα δεδομένα δεν μπορούν να χρησιμοποιηθούν για σκοπούς άλλους από αυτούς που δηλώθηκαν εξαρχής, όταν αποκτήθηκαν τα δεδομένα. Η εξόρυξη δεδομένων παραβιάζει δυνητικά και τις δύο αυτές αρχές.

## 6.4 Ακρίβεια των δεδομένων

Μία ακόμα κατευθυντήρια γραμμή του ΟΟΣΑ απαιτεί τα προσωπικά δεδομένα να είναι ακριβή, πλήρη και πρόσφατα, προκειμένου να προστατευτούν τα άτομα από τις επιβλαβείς συνέπειες που συνδέονται με την κακή ποιότητα των δεδομένων. Οι τεχνικές της εξόρυξης δεδομένων εφαρμόζονται σε τεράστιες ποσότητες δεδομένων, τα οποία είναι πιθανό να έχουν προέλθει από ποικίλες και πιθανόν εξωτερικές πηγές. Ως εκ τούτου, η αρχική ποιότητα των δεδομένων δεν μπορεί να είναι εξασφαλισμένη, με πιθανό αποτέλεσμα τα δεδομένα να είναι «θορυβώδη», παρωχημένα, ανακριβή και ελλιπή. Επιπλέον, παρόλο που τα δεδομένα τίθενται σε προ-επεξεργασία ή καθαρισμό πριν από την εκτέλεση μιας εφαρμογής εξόρυξης δεδομένων, τα άτομα διεξάγουν συναλλαγές σποραδικά και πολλές φορές με απρόβλεπτο τρόπο, ο οποίος μπορεί να προκαλέσει τη γρήγορη φθορά και ανακρίβεια των δεδομένων (Wahlstrom et al. 2002).

## 6.5 Ασφάλεια της βάσης δεδομένων

Η ασφάλεια της βάσης δεδομένων παρεμποδίζει την μη εξουσιοδοτημένη διάδοση των προσωπικών δεδομένων, και έτσι ενισχύει περαιτέρω, έστω και εμμέσως, την ικανότητα ενός ατόμου να ρυθμίζει την πρόσβαση στα δεδομένα του. Σύμφωνα με τους Chen, Han και Yu (1996), όταν τα δεδομένα μπορούν να προβληθούν από πολλές διαφορετικές οπτικές γωνίες και σε διαφορετικά αφαιρετικά επίπεδα, απειλείται ο στόχος της προστασίας της ασφάλειας των δεδομένων και της προστασίας ενάντια της εισβολής στην ιδιωτική ζωή. Είναι σημαντικό να γνωρίζουμε τους παράγοντες που μπορεί να οδηγήσουν στην παραβίαση της ιδιωτικής ζωής, καθώς και τα μέτρα ασφαλείας που μπορούν να αναπτυχθούν για την πρόληψη της αποκάλυψης ευαίσθητων πληροφοριών.

Η ανάπτυξη των αποθηκών δεδομένων έχει συνδράμει στην αυξανόμενη σημασία της ασφάλειας της βάσης δεδομένων. Προτού αναπτυχθούν οι αποθήκες δεδομένων, τα

δεδομένα κρατούνταν σε ξεχωριστές βάσεις δεδομένων, στις οποίες η πρόσβαση ήταν ελεγχόμενη και περιορισμένη σε άτομα με ένα συγκεκριμένο λειτουργικό ρόλο. Οι αποθήκες δεδομένων συγκεντρώνουν δεδομένα από πολλές πηγές και ως εκ τούτου, κατά τον καθορισμό των μέτρων ασφαλείας, πρέπει να λαμβάνονται υπόψη περισσότερο σύνθετοι παράγοντες (Wahlstrom et al. 2002).

## 6.6 Εξόρυξη δεδομένων & παραδοσιακές τεχνικές ανάκτησης πληροφοριών

Οι παραδοσιακές τεχνικές ανάκτησης πληροφοριών περιλαμβάνουν τη συγχώνευση των ηλεκτρονικών αρχείων από όλες τις βάσεις δεδομένων των υπολογιστών (computer merging) και την αντιστοίχιση των ηλεκτρονικών αρχείων από τις βάσεις δεδομένων (computer matching) (Fulda 1998). Τόσο η εξόρυξη δεδομένων, όσο και οι παραδοσιακές ηλεκτρονικές τεχνικές ανάκτησης πληροφοριών εξαρτώνται από τη χρήση μεγάλων βάσεων δεδομένων για την καταγραφή, αποθήκευση και ανταλλαγή προσωπικών πληροφοριών. Αν και η εξόρυξη δεδομένων μοιάζει να έχει πολλά κοινά με τις παραδοσιακές ηλεκτρονικές τεχνικές, οι ανησυχίες για την προστασία προσωπικών δεδομένων που προέρχονται από την εφαρμογή της εξόρυξης δεδομένων υπερβαίνουν κατά πολύ αυτές των παραδοσιακών τεχνικών, και οι λόγοι είναι οι ακόλουθοι (Tavani 1999).

1. Η διαδικασία της εξόρυξης δεδομένων παράγει έμμεσα μοτίβα και πρότυπα που περιλαμβάνουν πληροφορίες σχετικά με τα άτομα, κάτι το οποίο δε συμβαίνει στις παραδοσιακές τεχνικές ανάκτησης πληροφοριών, κατά τις οποίες χρησιμοποιούνται ρητά αρχεία δεδομένων για τη δημιουργία ενός συγχωνευμένου αρχείου.

2. Τόσο η ενσωμάτωση των ηλεκτρονικών αρχείων σε βάσεις δεδομένων ηλεκτρονικών υπολογιστών όσο και η αντιστοίχιση των ηλεκτρονικών αρχείων κατά βάσεις δεδομένων, συνεπάγονται την ανταλλαγή ρητών αρχείων που αφορούν περισσότερες από μία βάσεις δεδομένων. Σε αντίθεση, η εξόρυξη δεδομένων χρησιμοποιεί μία μεγάλη βάση δεδομένων, ονομαζόμενη και ως αποθήκη δεδομένων, και έτσι δεν είναι απαραίτητο να μεταφέρονται τα δεδομένα ή να ανταλλάσσονται με αρχεία σε εξωτερική βάση δεδομένων.

3. Στις παραδοσιακές τεχνικές χρησιμοποιούνται συγκεκριμένες ερωτήσεις για την ανάκτηση πληροφοριών και σχέσεων, οι οποίες είναι ήδη γνωστό ότι υπάρχουν. Παράδειγμα τέτοιων ερωτήσεων είναι «Πόσα προϊόντα συγκεκριμένου τύπου πούλησε η εταιρεία τη συγκεκριμένη χρονολογία και στη συγκεκριμένη περιοχή». Αντίθετα, η εξόρυξη δεδομένων χρησιμοποιεί ερωτήσεις ανοικτού τύπου και αλγόριθμους για να ανακαλύψει σχέσεις-κλειδιά των δεδομένων, οι οποίες ήταν άγνωστες προηγουμένως. Για παράδειγμα, ο χρήστης θα μπορούσε να διεξάγει ένα ερώτημα ή εντολή του τύπου: «Εμφάνισε όλα τα μοτίβα» ή «Εμφάνισε μία κατηγορία σχέσεων/τάσεων».

4. Οι εταιρείες που εφαρμόζουν την εξόρυξη δεδομένων δε μπορούν να γνωρίζουν εκ των προτέρων τι είδους δυνητικά πολύτιμα προσωπικά δεδομένα ή τι είδος σχέσεων μεταξύ των δεδομένων θα προκύψουν, ή το πού ακριβώς θα χρησιμοποιηθούν οι προκύπτουσες σχέσεις. Αντίθετα, οι παραδοσιακές τεχνικές παράγουν πληροφορίες με γενικά προγνωστική διάθεση. Μπορούν, για παράδειγμα, να προβλέψουν το πιθανό αποτέλεσμα των αντιστοιχισμένων αρχείων δεδομένων.

5. Οι περισσότερες πληροφορίες που παράγει η εξόρυξη δεδομένων είναι δημόσιου χαρακτήρα, ενώ οι παραδοσιακές τεχνικές ανάκτησης πληροφοριών ανακτούν και ανταλλάσσουν πληροφορίες ιδιωτικού χαρακτήρα.

6. Πολλές πρακτικές εξόρυξης δεδομένων παράγουν πληροφορίες σχετικά με μία συγκεκριμένη ομάδα ατόμων και όχι για τα ίδια τα άτομα, ενώ οι παραδοσιακές τεχνικές έχουν ως στόχο να ανακαλύψουν πληροφορίες για συγκεκριμένα άτομα τα οποία τυγχάνει να είναι μέλη προεπιλεγμένων ομάδων, και όχι πληροφορίες για τις ομάδες αυτές καθ’ αυτές.

## 6.7 Νέες απειλές των προσωπικών δεδομένων

Παρακάτω παρουσιάζονται οι νέες απειλές των προσωπικών δεδομένων που προκύπτουν από την εφαρμογή της εξόρυξης δεδομένων στα προσωπικά δεδομένα, επιβεβαιώνοντας ότι οι υφιστάμενοι νόμοι της ιδιωτικής ζωής δεν προσφέρουν επαρκή προστασία, λόγω των συνεχών εξελίξεων της τεχνολογίας. Η εξόρυξη δεδομένων χρησιμοποιεί τεράστιες ποσότητες δεδομένων για τη δημιουργία υποθέσεων και την

ανακάλυψη γενικών τάσεων. Συγκεκριμένα, η εξόρυξη δεδομένων θέτει τις ακόλουθες νέες προκλήσεις για την προστασία της ιδιωτικής ζωής (Brankovic and Estivill-Castro 1999):

- Στερεότυπα: Τα γενικά μοτίβα-μοντέλα ενδέχεται να οδηγήσουν σε προκαταλήψεις και στερεότυπα. Εάν τα μοτίβα βασίζονται σε χαρακτηριστικά όπως η φυλή, το φύλο ή η εθνικότητα, το θέμα αυτό μπορεί να γίνει ιδιαίτερα ευαίσθητο και αμφιλεγόμενο.

- Φύλαξη των προσωπικών δεδομένων από τους ερευνητές: Τα άτομα επωφελούνται από τις προσπάθειες συλλογής δεδομένων στη διαδικασία οικοδόμησης της γνώσης που οδηγεί την κοινωνία, όπως για παράδειγμα η κατανόηση των επιδημιών μέσω της ανάλυσης των ιατρικών αρχείων. Οι ερευνητές θεωρούν ότι οι διατάξεις περί απορρήτου θα επιβάλουν τόσους πολλούς περιορισμούς στα δεδομένα, που θα τα καταστήσει άχρηστα. Έτσι, θα πρέπει να βρεθεί η κατάλληλη ισορροπία ανάμεσα στην ανάγκη για προστασία της ιδιωτικής ζωής και την ανάγκη για άντληση γνώσεων.

- Συνδυασμός των μοτίβων-προτύπων: Ο συνδυασμός δύο ή περισσότερων γενικών τάσεων μπορεί να οδηγήσει σε αποκάλυψη των ατομικών στοιχείων, είτε με βεβαιότητα, είτε με υψηλή πιθανότητα.

## 6.8 Τεχνικές για την προστασία των προσωπικών δεδομένων στην εξόρυξη δεδομένων

Η προστασία της προσωπικής ζωής και τα ηθικά ζητήματα έχουν λάβει σχετικά μικρό ενδιαφέρον στη διαδικασία της εξόρυξης δεδομένων. Αντίθετα, στις βάσεις στατιστικών δεδομένων χρησιμοποιούνται ορισμένες τεχνικές για την ενίσχυση της προστασίας των προσωπικών δεδομένων. Οι τεχνικές αυτές θα μπορούσαν να εφαρμοστούν και στη διαδικασία της εξόρυξης δεδομένων. Οι τεχνικές αυτές είναι οι ακόλουθες (Roddick and Fule 2003):

- Η ανωνυμία των προσωπικών δεδομένων (anonymisation), στα οποία, οποιοδήποτε προσδιοριστικό χαρακτηριστικό αφαιρείται από το σύνολο των δεδομένων. Μία

παραλλαγή αυτού μπορεί να είναι ένα φίλτρο, το οποίο εφαρμόζεται στο σύνολο των κανόνων για την καταστολή των κανόνων που περιέχουν προσδιοριστικά χαρακτηριστικά.

- Περιορισμοί ερωτημάτων (query restriction), οι οποίοι προσπαθούν να ανιχνεύσουν πότε οι στατιστικές λύσεις μπορεί να είναι δυνατές μέσω συνδυασμών ερωτήσεων. Σύμφωνα με τους Brankovic και Estivill-Castro (1999), οι τεχνικές περιορισμού ερωτημάτων παρέχουν ακριβείς απαντήσεις σε ορισμένα ερωτήματα και απορρίπτουν άλλα ερωτήματα, τα οποία μπορεί να οδηγήσουν σε συμβιβαστικές λύσεις της βάσης δεδομένων.

- Δυναμική δειγματοληψία (dynamic sampling) και μείωση του μεγέθους των διαθέσιμων δεδομένων ή του δείγματος δεδομένων (reducing the size of available data). Αυτό μπορεί να επιτευχθεί επιλέγοντας ένα διαφορετικό σύνολο πηγών δεδομένων για κάθε ερώτημα.

- Προσθήκη θορύβου και διατάραξη των δεδομένων από μεμονωμένες καταχωρήσεις με τέτοιο τρόπο ώστε να διατηρείται η ακρίβεια των στατιστικών ερωτημάτων. Αυτό μπορεί να επιτευχθεί με τους ακόλουθους δύο τρόπους:

  - ✓ Προσθήκη θορύβου (noise addition), κατά την οποία τα σύνολα τιμών μεταβάλλονται κατά τέτοιο τρόπο, ώστε οι κοινές στατιστικές και εξορυκτικές δραστηριότητες να οδηγούν στο ίδιο αποτέλεσμα.

  - ✓ Εναλλαγή δεδομένων (data swapping), η οποία αποτελεί μία μορφή προσθήκης θορύβου και κατά την οποία οι τιμές στα αρχεία εναλλάσσονται με τέτοιο τρόπο ώστε να διαφυλάσσονται τα στατιστικά στοιχεία. Αυτές οι τεχνικές, σε συνδυασμό με τη δειγματοληψία, δημιουργούν ένα συνδυασμό ακρίβειας και ιδιωτικότητας (Evfimievski et al. 2004).

Οι Agrawal και Srikant στο άρθρο τους (2000) διερευνούν τη δυνατότητα διατήρησης της προστασίας των προσωπικών δεδομένων στην εξόρυξη δεδομένων και παρουσιάζουν τις ακόλουθες δύο τεχνικές:

- Συμμετοχή κατηγοριών τιμών (value-class membership), κατά την οποία οι τιμές για ένα χαρακτηριστικό είναι χωρισμένες σε ένα σύνολο ξένων αλληλοαποκλειόμενων τάξεων.

- Παραμόρφωση τιμών (value distortion), η οποία επιστρέφει τη διατάραξη μίας τιμής x + r, αντί της x, όπου r είναι μία τυχαία τιμή, η οποία μπορεί να υπολογιστεί είτε από μία ομοιόμορφη κατανομή, είτε από μία τυχαία κατανομή Gaussian.

# ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ

## 7.1 Συμπεράσματα

Με την αυτόματη ανακάλυψη γνώσης μέσα από τις βάσεις δεδομένων, η εξόρυξη δεδομένων χρησιμοποιεί εξελιγμένες στατιστικές αναλύσεις και τεχνικές μοντελοποίησης για να αποκαλύψει κρυμμένα πρότυπα, κανόνες και σχέσεις σε οργανωμένες βάσεις δεδομένων. Κατά την διάρκεια των τελευταίων 40 ετών, τα εργαλεία και οι τεχνικές επεξεργασίας δομημένων πληροφοριών έχουν συνεχίσει να εξελίσσονται από τις βάσεις δεδομένων στις αποθήκες δεδομένων. Οι εφαρμογές στις αποθήκες δεδομένων έχουν καταστεί κρίσιμες για τις επιχειρηματικές δραστηριότητες.

Παρότι η εξόρυξη δεδομένων βρίσκεται ακόμα σε αρχικό στάδιο, ένα ευρύ φάσμα εταιρειών-βιομηχανιών, συμπεριλαμβανομένων των εταιρειών λιανικής πώλησης, των τραπεζών, των εταιρειών υγειονομικού ενδιαφέροντος, των τηλεπικοινωνιών, των χρηματοδοτικών εταιρειών κλπ, χρησιμοποιούν ήδη εργαλεία και τεχνικές εξόρυξης δεδομένων για να επωφεληθούν από τα ιστορικά τους στοιχεία. Με τη χρήση των τεχνολογιών αναγνώρισης-προτύπων αλλά και στατιστικές και μαθηματικές τεχνικές, η εξόρυξη δεδομένων βοηθάει τους αναλυτές να εντοπίζουν σημαντικά γεγονότα, σχέσεις, πρότυπα, εξαιρέσεις αλλά και ανωμαλίες, οι οποίες θα μπορούσαν να περάσουν απαρατήρητες. Ιδιαίτερα χρήσιμη είναι η εφαρμογή τους και στις αποφάσεις μάρκετινγκ σε θέματα όπως η διαχείριση πελατειακών σχέσεων, το διαδραστικό μάρκετινγκ σε πραγματικό χρόνο και τη δημιουργία του προφίλ των πελατών. Για το λόγο αυτό, υπάρχει η ανάγκη για βαθύτερη κατανόηση της χρήσης της εξόρυξης

δεδομένων και διαχείρισης της γνώσης για την υποστήριξη των αποφάσεων μάρκετινγκ.

Η εξόρυξη δεδομένων έχει εξελιχθεί σε μεγάλο βαθμό για διάφορους λόγους: οι οργανισμοί συγκεντρώνουν περισσότερες πληροφορίες για τις επιχειρηματικές τους δραστηριότητες, το κόστος της αποθήκευσης δεδομένων έχει μειωθεί δραστικά και οι ανταγωνιστικές πιέσεις των επιχειρήσεων έχουν αυξηθεί. Άλλοι παράγοντες είναι η εμφάνιση πιέσεων για τον έλεγχο των υφιστάμενων επενδύσεων, αλλά και η ένδειξη του δείκτη κόστους/απόδοσης των συστημάτων των ηλεκτρονικών υπολογιστών. Οι εφαρμογές της εξόρυξης δεδομένων είναι πολλές: από την ανάλυση και πρόβλεψη της συμπεριφοράς των πελατών, μέχρι και το σχεδιασμό στοχευμένων εκστρατειών μάρκετινγκ. Τέτοιες εφαρμογές είναι στενά συνδεδεμένες με τη βελτιστοποίηση του κύκλου ζωής των πελατών και την αποτελεσματική διαχείριση πελατειακών σχέσεων.

Η διαχείριση πελατειακών σχέσεων από τις επιχειρήσεις κρίνεται απαραίτητη, προκειμένου να παραμείνουν ανταγωνιστικές στη σημερινή αγορά. Η αποτελεσματική χρήση των πληροφοριών των πελατών συνεπάγεται την άμεση ανταπόκριση στις ανάγκες τους, και κατά συνέπεια στην κερδοφορία. Προϋπόθεση για μία επιτυχημένη επιχείρηση είναι η κατανόηση των πελατών και των αναγκών τους, και η εξόρυξη δεδομένων αποτελεί το μέσο-κλειδί.

Υπάρχουν τέσσερις βασικές λειτουργίες της εξόρυξης δεδομένων, που υποστηρίζονται από πολυάριθμες τεχνικές εξόρυξης:

1. Δημιουργία μοντέλου πρόβλεψης, που υποστηρίζεται από επαγωγικές τεχνικές,
2. Ανάλυση σχέσεων, που υποστηρίζεται από τεχνικές ανακάλυψης ακολουθιών,
3. Κατάτμηση της βάσης δεδομένων, που υποστηρίζεται από τεχνικές ομαδοποίησης-συσπείρωσης, και
4. Ανίχνευση αποκλίσεων, η οποία υποστηρίζεται από στατιστικές τεχνικές.

Η ραγδαία εξάπλωση του παγκοσμίου ιστού τα τελευταία χρόνια επιτείνει την ανάγκη ανάπτυξης συστημάτων που να βοηθούν τους χρήστες να εκμεταλλευτούν τον τεράστιο όγκο κειμένου και πληροφοριών που είναι διαθέσιμος στο διαδίκτυο. Το web mining έχει ως στόχο τη μελέτη των εφαρμογών της εξόρυξης δεδομένων στα δεδομένα του διαδίκτυο. Το διαδίκτυο αποτελεί την καλύτερη βάση για την εφαρμογή της εξόρυξης δεδομένων, η οποία είναι σε θέση να βοηθήσει τον διαχειριστή της ιστοσελίδας να

προσδιορίσει το προφίλ των πελατών, να αναλύσει τις συμπεριφορές τους και να έχει άμεση ανατροφοδότηση, ώστε να δημοσιεύονται εξατομικευμένες ηλεκτρονικές πληροφορίες και εν τέλει, να παρέχεται η καλύτερη δυνατή διαχείριση των πελατειακών σχέσεων.

Η ανάλυση των δεδομένων των χρηστών, μέσω του web mining, αποκαλύπτει σε πραγματικό χρόνο τις γεωγραφικές περιοχές στις οποίες η εταιρεία έχει ευκαιρίες για το ηλεκτρονικό επιχειρείν, παρέχει τρόπους στόχευσης και κατανόησης των αναγκών των πελατών, ώστε να προσαρμόσουν τις υπηρεσίες και στρατηγικές τους σε πραγματικό χρόνο, και συνεισφέρει στην αποτελεσματικότητα ενός δικτυακού τόπου ως κανάλι του μάρκετινγκ, ποσοτικοποιώντας τις συμπεριφορές του χρήστη όσο βρίσκεται στην ιστοσελίδα.

Η τουριστική βιομηχανία αποτελεί έναν από τους μεγαλύτερους χρήστες της τεχνολογίας. Με την εξόρυξη γνώσης, χρησιμοποιώντας αποθηκευμένες και λεπτομερείς πληροφορίες για τους πελάτες, την αγορά και τις υπηρεσίες, οι τουριστικοί οργανισμοί και τουριστικοί πράκτορες είναι κατάλληλα εξοπλισμένοι ώστε να συμβαδίζουν και να αντιμετωπίζουν τις συνεχώς μεταβαλλόμενες ανάγκες της αγοράς. Οι τεχνικές της εξόρυξης δεδομένων μπορούν να εφαρμοστούν στην τουριστική βιομηχανία, και συγκεκριμένα στο διαδικτυακό μάρκετινγκ της τουριστικής βιομηχανίας, προσφέροντας πολλά πλεονεκτήματα. Μέσω του μοντέλου των Wong, Chen, Chung και Kao (2006), ο τουριστικός οργανισμός είναι σε θέση να ανακαλύψει τα χαρακτηριστικά των πολύτιμων πελατών και τους προτιμώμενους επόμενους προορισμούς ταξιδίων. Έπειτα, είναι σε θέση να τμηματοποιήσει τους ταξιδιώτες για τους σκοπούς των σταυροειδών πωλήσεων.

Οι τεράστιες ποσότητες δεδομένων που προκύπτουν από τις συναλλαγές στην υγειονομική περίθαλψη αποδεικνύονται υπερβολικά πολύπλοκες και ογκώδεις για να επεξεργαστούν και να αναλυθούν από τις παραδοσιακές μεθόδους. Η εξόρυξη δεδομένων παρέχει την απαραίτητη μεθοδολογία και την τεχνολογία για να μετατρέψει τους τεράστιους αυτούς όγκους δεδομένων σε χρήσιμες πληροφορίες για τη λήψη αποφάσεων. Υπάρχουν τρία επίπεδα εφαρμογών της εξόρυξης δεδομένων στη βιομηχανία της υγείας, το χαμηλό επίπεδο (low level), το νοσοκομειακό επίπεδο (hospital-based) και το παγκόσμιο επίπεδο (global level). Η εφαρμογή της εξόρυξης

δεδομένων στη διαχείριση πελατειακών σχέσεων αναφέρεται στο νοσοκομειακό επίπεδο.

Τα εργαλεία της εξόρυξης δεδομένων μπορούν να εφαρμοστούν σε διάφορους επιστημονικούς και επιχειρηματικούς τομείς, ανάμεσά τους και ο τραπεζικός τομέας, το λιανικό εμπόριο και ο κλάδος των τηλεπικοινωνιών. Οι εφαρμογές της εξόρυξης στον τραπεζικό τομέα συμβάλλουν στην αποτελεσματική βαθμολόγηση της πιστοληπτικής ικανότητας των πελατών, την έγκριση των τραπεζικών δανείων, την πρόβλεψη αθέτησης πληρωμής, την ανίχνευση παράνομων συναλλαγών, την διαχείριση των διαθεσίμων και τη βελτιστοποίηση των χαρτοφυλακίων μετοχών. Στο λιανικό εμπόριο, η εξόρυξη δεδομένων μπορεί να χρησιμοποιηθεί στην πρόβλεψη των πωλήσεων της επιχείρησης, τις εποχιακές διακυμάνσεις, ή ακόμα και στην πρόβλεψη των γενικών τάσεων της αγοράς. Επίσης, είναι δυνατή η παροχή πληροφοριών σχετικά με τις αγοραστικές συνήθειες και τις προτιμήσεις των πελατών. Τέλος, η εξόρυξη δεδομένων στον κλάδο των τηλεπικοινωνιών εφαρμόζεται σε τρεις κύριους τομείς, οι οποίοι είναι η ανίχνευση της απάτης, η δημιουργία προφίλ πελατών και η απομόνωση βλαβών του δικτύου.

Η εξόρυξη δεδομένων, παρά τη σημαντική συνεισφορά της στην ανακάλυψη της γνώσης, αποτελεί απειλή για ορισμένες σημαντικές ηθικές αξίες, όπως η προστασία των προσωπικών δεδομένων. Η προστασία προσωπικών δεδομένων συνήθως θεωρείται ως το δικαίωμα του ατόμου για τον έλεγχο των πληροφοριών που αφορούν τον εαυτό του. Οι υφιστάμενοι νόμοι περί προστασίας της ιδιωτικής ζωής έχουν απαρχαιωθεί από τις εξελίξεις τις τεχνολογίας και δεν προσφέρουν πλέον επαρκή προστασία. Για το λόγο αυτό, υπάρχει μία αυξανόμενη ανησυχία των πολιτών για την προστασία της προσωπικής ζωής των ατόμων.

Το 1980 ο Οργανισμός για Οικονομική Συνεργασία και Ανάπτυξη (ΟΟΣΑ – OECD) ενέκρινε τις κατευθυντήριες γραμμές για την προστασία των προσωπικών δεδομένων. Δύο από αυτές τις κατευθυντήριες γραμμές ορίζουν ότι ο λόγος για τη συλλογή προσωπικών δεδομένων θα πρέπει να καταστεί γνωστός και σαφής στο άτομο πριν από τη συλλογή τους και ότι τα δεδομένα δεν μπορούν να χρησιμοποιηθούν για σκοπούς άλλους από αυτούς που δηλώθηκαν εξαρχής, όταν αποκτήθηκαν τα δεδομένα. Η εξόρυξη δεδομένων παραβιάζει δυνητικά και τις δύο αυτές αρχές. Για την ενίσχυση της προστασίας των προσωπικών δεδομένων, θα μπορούσαν να εφαρμοστούν στη

διαδικασία της εξόρυξης ορισμένες τεχνικές, οι οποίες χρησιμοποιούνται στις βάσεις στατιστικών δεδομένων, όπως η ανωνυμία των προσωπικών δεδομένων, η προσθήκη θορύβου και η παραμόρφωση τιμών.

Η εξόρυξη δεδομένων αποτελεί ένα πολύ ισχυρό εργαλείο που θα πρέπει να χρησιμοποιείται με τη μέγιστη προσοχή για την αύξηση της ικανοποίησης των πελατών και την παροχή ποιοτικότερων και ασφαλέστερων προϊόντων, με τρόπο που δεν επηρεάζει την ιδιωτική ζωή και το ιδιωτικό απόρρητο. Δε θα ήταν υπερβολικά αισιόδοξο αν λέγαμε ότι η εξόρυξη δεδομένων έχει ένα πολλά υποσχόμενα μέλλον και ότι τα επόμενα χρόνια θα φέρουν πολλές νέες εξελίξεις, μεθόδους και τεχνολογικές καινοτομίες. Επιπλέον, η βελτιωμένη ενοποίηση των τεχνικών εξόρυξης δεδομένων μπορεί να επιφέρει τη χρησιμοποίηση νέων τύπων δεδομένων και εφαρμογών.

## 7.2. Περιορισμοί και προτάσεις για περαιτέρω έρευνα

Στα πλαίσια της παρούσας εργασίας κατέστη δυνατό να διερευνηθούν δύο κλάδοι εφαρμογής της εξόρυξης δεδομένων, η τουριστική βιομηχανία και ο τομέας της υγείας, ενώ αναφέρθηκαν περιληπτικά οι εφαρμογές της στον τραπεζικό τομέα, το λιανικό εμπόριο και τον κλάδο των τηλεπικοινωνιών. Κατά συνέπεια, θα πρέπει να πραγματοποιηθεί περαιτέρω έρευνα για την εφαρμογή των τεχνικών εξόρυξης δεδομένων σε περισσότερους κλάδους. Η ανάλυση των εφαρμογών στον κλάδο της υγείας περιορίστηκε στο νοσοκομειακό επίπεδο, ενώ η ανάλυση στη βιομηχανία του τουρισμού οριοθετήθηκε στον τουριστικό οργανισμό.

Σε βιβλιογραφικό επίπεδο, λόγω του περιορισμένου μεγέθους της συγκεκριμένης εργασίας, η μελέτη περιορίζεται στον αριθμό των ακαδημαϊκών πηγών, σε σχέση με τις έννοιες των όρων, τις μεθόδους και τις τεχνικές που αναφέρονται στην παρούσα μελέτη. Για το λόγο αυτό, επιπρόσθετη έρευνα κρίνεται αναγκαία, προκειμένου να επιτευχθεί η σε βάθος κατανόηση όλων των πτυχών της εργασίας.

Η εξόρυξη δεδομένων, ως ένας σχετικά νέος και πολλά υποσχόμενος τομέας, εξακολουθεί να αντιμετωπίζει πολλές προκλήσεις και ανεπίλυτα προβλήματα που θέτουν νέα ερευνητικά ζητήματα για περαιτέρω μελέτη. Τέτοια θέματα αποτελούν η ανάπτυξη μίας ευέλικτης και κατανοητής γλώσσας απόδοσης των αποτελεσμάτων, η

ανάπτυξη τεχνικών εξόρυξης σε προηγμένα συστήματα βάσεων δεδομένων, όπως οι ενεργές βάσεις δεδομένων (active databases) και οι χωρικές βάσεις δεδομένων (spatial databases), η ενσωμάτωση της ανακαλυφθείσας γνώσης με τη γνώση των εμπειρογνωμόνων, και η ανάπτυξη μεθόδων για τη διασφάλιση της ασφάλειας και την προστασία της ιδιωτικής ζωής.

Με τη μετάβαση από το μαζικό μάρκετινγκ (mass marketing) στο εξατομικευμένο μάρκετινγκ (one-to-one marketing), και η περαιτέρω ανάπτυξη των εργαλείων εξόρυξης δεδομένων μπορεί να ωφελήσει όλες τις λειτουργίες του μάρκετινγκ. Η συστηματική εφαρμογή των τεχνικών εξόρυξης δεδομένων θα ενισχύσει τη διαδικασία της διαχείρισης γνώσης και θα εφοδιάσει την επιχείρηση με τις απαραίτητες γνώσεις για την αναγνώριση των αναγκών των πελατών και την καλύτερη εξυπηρέτησή τους. Η απόκτηση δεδομένων, για τη βαθύτερη κατανόηση των πελατών, θα πρέπει να γίνεται με μη-παρεμβατικό τρόπο, με χαμηλό κόστος, και υψηλή ακρίβεια. Επειδή η συμπεριφορά των πελατών είναι αβέβαιη και, πολλές φορές, αντιφατική, οι ερευνητές θα πρέπει να κατασκευάσουν ένα δυναμικό μοντέλο τμηματοποίησης πελατών, το οποίο θα αντανακλά αυτά τα χαρακτηριστικά.

Τα διαδικτυακά δεδομένα αυξάνονται με σημαντικό ρυθμό, γεγονός το οποίο καθιστά το web mining μια εύφορη περιοχή περαιτέρω έρευνας. Η τεχνολογία του παγκόσμιου ιστού έχει μεγάλο αντίκτυπο στη διαχείριση γνώσης και την εφαρμογή της εξόρυξης δεδομένων, και οι μελλοντικές εξελίξεις των συστημάτων πληροφοριών θα επιφέρουν ακόμα μεγαλύτερη πρόοδο στον τομέα του web mining.

## ΠΑΡΑΡΤΗΜΑ

Πίνακας 1: Κατηγοριοποίηση των πηγών στα κεφάλαια της εργασίας.

| ΚΕΦΑΛΑΙΟ | ΣΥΓΓΡΑΦΕΙΣ – ΤΙΤΛΟΣ ΑΡΘΡΟΥ |
|---|---|
| **1. Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ** | Ahmed, S. R. (2004) Application of data mining in Retail Business.<br><br>Berry, N.J.A. and Linoff, G.S. (1999) Mastering Data Mining: The Art and Science of Customer Relationship Management.<br><br>Chen, M. S., Han, J., and Yu, P. S. (1996) Data Mining: An Overview From Database Perspective.<br><br>Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996) From Data Mining to Knowledge Discovery in Databases.<br><br>Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996) The KDD Process for Extracting Useful Knowledge from Volumes of Data.<br><br>Frawley, W. J., Piatetsky – Shapiro G., and Matheus, C. J. (1992) Knowledge Discovery In Databases.<br><br>Shaw, M. J., Subramaniam, C., Tan, G. W., and Welge, M. E. (2001) Knowledge Management and Data Mining for Marketing.<br><br>Wu, X., Yu, P. S., Piatetsky-Shapiro, G., Cercone, N., Lin, T. Y., Kotagiri, R., and Wah, B. W. (2003) Data Mining: How Research Meets Practical Development? |
| **2. Η ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ (CRM)** | Ahn, J. Y., Kim, S. K., and Han, K. S. (2003) On the Design Concepts for CRM Systems.<br><br>Arndt, D. and Gestern, W. (2001). Data Management in Analytical Customer Relationship Management.<br><br>Bose, R. (2002) Customer Relationship Management: Key Components for IT Success. |

| | |
|---|---|
| | Bounsaythip, C., and Rinta-Runsala, E. (2001) Overview of Data Mining for Customer Behavior Modeling.<br><br>Goyat. S. (2011) The Basis of Market Segmentation: A Critical Review of Literature.<br><br>He, Z., Xu, X., Huang, J. Z., and Deng, S. (2004) Mining Class Outliers: Concepts, Algorithms and Applications in CRM.<br><br>Liu, C. N., and Zhu, X. W. (2009) A Study on CRM Technology Implementation and Application Practices.<br><br>Parvatiyar, A., and Sheth, J. N. (2001) Customer Relationship Management: Emerging Practice, Process, and Discipline.<br><br>Wilson, R. F. (2000) Preparing a Customer Profile for Your Internet Marketing Plan.<br><br>Xu, M., and Walton, J. (2005) Gaining Customer Knowledge through Analytical CRM. |
| **3. Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ** | Chen, Y., Zhang, G., Hu, D., and Wang, S. (2006) Customer Segmentation in Customer Relationship Management Based on Data Mining.<br><br>Edelstein, H. (2000) Building Profitable Customer Relationships with Data Mining, executive briefing.<br><br>Eriksson, K., and Vaghult, A. L. (2000) Customer Retention, Purchasing Behavior and Relationship Substance in Professional Services.<br><br>Hosseini, S. M. S., Maleki, A., and Gholamian, M. R. (2010) Cluster Analysis Using Data Mining Approach to Develop CRM Methodology to Assess the Customer Loyalty.<br><br>Kimball, R. (1997) Preparing For Data Mining. |

| | |
|---|---|
| | Ngai, E. W. T., Xiu, L., and Chau, D. C. K. (2009) Application of Data mining Techniques in Customer Relationship Management: A literature review and classification.<br><br>Payne, A. (2002) The value creation process in customer relationship management.<br><br>Rygielski, C., Wang, J. C., and Yen, D. C. (2002) Data Mining Techniques for Customer Relationship Management.<br><br>Sohrabi, B., and Khanlari, A. (2007) Customer Lifetime Value Measurement Based on RFM Model.<br><br>Thearling, K. (2000) Data Mining and Customer Relationships.<br><br>Thearling, K. (2010) Data Mining for CRM.<br><br>Yang, A. X. (2004) How to develop new approaches to RFM segmentation. |
| **4. ΤΟ WEB MINING ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ**. | Araya, S., Silva, M., and Weber, R. (2004) A Methodology for Web Usage Mining and its Application to Target Group Identification.<br><br>Chou, P. H., Li, P. H., Chen, K. K., and Wu, M. J. (2010) Integrating Web Mining and Neural Network for Personalized e-Commerce Automatic Service.<br><br>Cooley, R., Mobasher, B., and Srivastava, J. (1997) Web Mining: Information and Pattern Discovey on the World Wide Web.<br><br>Huang, S. M., Kwan, I., and Li, S. H. (2003) Web Mining for CRM – An Empirical Study of Computer Game Service Company.<br><br>Madria, S. K., Bhowmick, S. S., Ng, W. K., and Lim, E. P. (1999) Research Issues in Web Data Mining.<br><br>Mobasher, B., Dai, H., Luo, T., and Nakagawa, M. (2002) Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization. |

| | |
|---|---|
| | Mobasher, B., Jain, N., Han, E. H., and Srivastava, J. (1996) Web Mining: Pattern Discovery from World wide Web Transactions.<br><br>Nasraoui, O., Soliman, M., Saka, E., Badia, A., and Germain, R. (2008) A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites.<br><br>Purandare, P. (2008) Web Mining: A Key to Improve Business on Web.<br><br>Srivastava, J., Cooley, R. W., Deshpande, M., Tan, P. N. (2000), Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. |
| **5. ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΠΕΛΑΤΕΙΑΚΩΝ ΣΧΕΣΕΩΝ** | Biafore, S. (1999) Predictive Solutions Bring More Power to Decision Makers.<br><br>Buhalis, D. (1999) Information Technology for Small and Medium-Sized Tourism Enterprises: Adaptation and Benefits.<br><br>Hallick, J. N. (2001) Analytics and the Data Warehouse.<br><br>Hsieh, N. - C. (2004) An Integrated Data Mining and Behavioral Scoring Model for Analyzing Bank Customers.<br><br>Koh, H. C., and Tan, G. (2005) Data Mining Applications in Healthcare.<br><br>Magnini, V. P., Honeycutt Jr., E. D., and Hodge, S. K. (2003) Data Mining for Hotel Firms: Use and Limitations.<br><br>Olmeda, I., and Sheldon, P. J. (2001) Data Mining Techniques and Applications for Tourism Internet Marketing.<br><br>Pyo, S., Uysal, M., and Chang, H. (2002) Knowledge Discovery in Database for Tourist Destinations.<br><br>Thanuja, V., Venkateswarlu, B., and Anjanjeyulu, G. S. G. N. (2011) Applications of Data Mining in Customer Relationship Management. |

| | |
|---|---|
| | Viswanathan, M. (2007) Distributed Data Mining in a Ubiquitous Healthcare Framework.<br><br>Watson, R.T., G.M. Zinkhan and L.F. Pitt (2000). Integrated Internet Marketing.<br><br>Weiss, G. M. (2005) Data Mining in Telecommunications.<br><br>Wong, J. Y., Chen, H. J., Chung, P. H., and Kao, N. C. (2006) Identifying Valuable Travelers and Their Next Foreign Destination by the Application of Data Mining Techniques.<br><br>Wyner. G. A. (1996) Customer Valuation: Linking Behavior and Economics.<br><br>Yina, W. (2010) Application of Customer Relationship Management in Health Care. |
| **6. ΗΘΙΚΑ ΖΗΤΗΜΑΤΑ ΣΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ** | Agrawal R. and Srikant, R. (2000) Privacy – Preserving Data Mining.<br><br>Birnhack, M. D. (2008) The EU Data Protection Directive: An Engine of a Global Regime.<br><br>Brankovic, L. and Estivill-Castro, V. (1999) Privacy Issues in Knowledge Discovery and Data Mining.<br><br>Culnan, M. J. (1993) How Did They Get My Name?: An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use.<br><br>Evfimievski, A., Srikant, R., Agrawal, R. and Gehrke, J. (2004) Privacy Preserving Mining of Association Rules.<br><br>Fulda, J. S. (1998) Data Mining and theWeb.<br><br>Laudon, K. C. (1996) Markets and Privacy.<br><br>OECD (1980) Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data. |

| | |
|---|---|
| | Roddick, J. F. and Fule, P. (2003) A System for Detecting Privacy and Ethical Sensitivity in Data Mining Results.<br><br>Tavani, H. T. (1999) Informational Privacy, Data Mining, and the Internet.<br><br>Wahlstrom, K., Roddick, J. F., Sarre, R., Estivill-Castro, V. and de Vries, D. (2002), On the Ethics of Data Mining. |

# ΒΙΒΛΙΟΓΡΑΦΙΑ


Agrawal R. and Srikant, R. (2000) Privacy – Preserving Data Mining. Chen, W., Naughton, J., and Bernstein, P. A. In: *Proceedings of the ACM SIGMOD Conference on the Management of Data,* Dallas, TX, ACM Press, pp. 439–450.

Ahmed, S. R. (2004) Application of data mining in Retail Business. In*: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04).* IEEE Computer Society Washington, DC, USA 2044.

Ahn, J. Y., Kim, S. K., and Han, K. S. (2003) On the Design Concepts for CRM Systems. *Industrial Management and Data Systems, Vol.* 103, No. 5, pp. 324-331.

Araya, S., Silva, M., and Weber, R. (2004) A Methodology for Web Usage Mining and its Application to Target Group Identification. *Fuzzy Sets and Systems,* Vol. 148, pp. 139-152.

Arndt, D., and Gestern, W. (2001). Data Management in Analytical Customer Relationship Management. *Data Mining for Marketing Applications Workshop at ECML/PKDD.*

Berry, N. J. A., and Linoff, G. S. (1999) Mastering Data Mining: The Art and Science of Customer Relationship Management 1st ed. New York, John Wiley & Sons, Inc.

Biafore, S. (1999) Predictive Solutions Bring More Power to Decision Makers. *Health Management Technology,* Vol. 20, No. 10, pp. 12-14.

Birnhack, M. D. (2008) The EU Data Protection Directive: An Engine of a Global Regime. *Computer Law and Security Report,* Vol. 24, pp. 508-520.

Bose, R. (2002) Customer Relationship Management: Key Components for IT Success. *Industrial Management and Data Systems,* Vol. 102, No. 2, pp. 89-97.

Bounsaythip, C., and Rinta-Runsala, E. (2001) Overview of Data Mining for Customer Behavior Modeling. Research report TTE1-2001-18, VTT Information Technology.

Brankovic, L. and Estivill-Castro, V. (1999) Privacy Issues in Knowledge Discovery and Data Mining. In: *Australian Institute of Computer Ethics Conference (AICEC99),* Melbourne, Australia, pp. 89-99.

Buhalis, D. (1999) Information Technology for Small and Medium-Sized Tourism Enterprises: Adaptation and Benefits. *Journal of Information Technology and Tourism.* Vol. 2, No.2, pp. 79-95.

Chen, M. S., Han, J., and Yu, P. S. (1996) Data Mining: An Overview from Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp. 866-883.

Chen, Y., Zhang, G., Hu, D., and Wang, S. (2006) Customer Segmentation in Customer Relationship Management Based on Data Mining. In: *International Federation for Information Processing (IFIP),* Knowledge Enterprise: Intelligent Strategies In Product Design, Manufacturing and Management eds. K. Wang, Kovacs G., Wozny M., Fang M., (Boston: Springer), Vol. 207, pp. 288-293.

Chou, P. H., Li, P. H., Chen, K. K., and Wu, M. J. (2010) Integrating Web Mining and Neural Network for Personalized e-Commerce Automatic Service. *Expert Systems with Applications,* Vol. 37, pp. 2898-2910.

Cooley, R., Mobasher, B., and Srivastava, J. (1997) Web Mining: Information and Pattern Discovey on the World Wide Web. In: *Proceedings of Ninth IEEE International Conference on Tools with Artificial Intelligence*, Newport Beach, CA, November 1997, pp. 558-567.

Culnan, M. J. (1993) How Did They Get My Name?: An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS Quarterly*, Vol. 17, No. 3, pp. 341-363.

Edelstein, H. (2000) Building Profitable Customer Relationships with Data Mining, executive briefing. SPSS Inc, Two Crows Corporation.

Evfimievski, A., Srikant, R., Agrawal, R. and Gehrke, J. (2004) Privacy Preserving Mining of Association Rules. *Information Systems,* Vol. 29, No. 4, pp. 343-364.

Eriksson, K., and Vaghult, A. L. (2000) Customer Retention, Purchasing Behavior and Relationship Substance in Professional Services. *Industrial Marketing Management,* Vol. 29, pp. 363-372.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996) The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Comunication of the ACM,* Vol. 39, No. 11, pp. 27-34.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996) From Data Mining to Knowledge Discovery in Databases. *AI Magazine,* Vol. 17, pp. 37-54.

Frawley, W. J., Piatetsky-Shapiro G., and Matheus, C. J. (1992) Knowledge Discovery In Databases, *AAAI / MIT Press*.

Fulda, J. S. (1998) Data Mining and the Web. *Computers and Society,* Vol. 28, No. 1, pp. 42–43.

Goyat. S. (2011) The Basis of Market Segmentation: A Critical Review of Literature. *European Journal of Business and Management,* Vol 3, No.9, pp. 45-54.

Hallick, J. N. (2001) Analytics and the Data Warehouse. *Health Management Technology*, Vol. 22, No. 6, pp. 24-25.

He, Z., Xu, X., Huang, J. Z., and Deng, S. (2004) Mining Class Outliers: Concepts, Algorithms and Applications in CRM. *Expert Systems with Applications,* Vo. 27, pp. 681-697.

Hosseini, S. M. S., Maleki, A., and Gholamian, M. R. (2010) Cluster Analysis Using Data Mining Approach to Develop CRM Methodology to Assess the Customer Loyalty. *Expert Systems with Applications,* Vol. 37, pp. 5259-5264.

Hsieh, N. - C. (2004) An Integrated Data Mining and Behavioral Scoring Model for Analyzing Bank Customers. *Expert Systems with Applications,* Vol. 27, pp. 623-633.

Huang, S. M., Kwan, I., and Li, S. H. (2003) Web Mining for CRM – An Empirical Study of Computer Game Service Company. *International Journal Of Electronic Business Management,* Vol. 1, No. 1, pp. 36-45.

Kimball, R. (1997) Preparing For Data Mining. *DBMS Online,* Nov. 1997. Available from: http://www.kimballgroup.com/html/articles_search/articles1997/9711d05.html.

Koh, H. C., and Tan, G. (2005) Data Mining Applications in Healthcare. *Journal of Healthcare Information Management,* Vol. 19, No. 2, pp. 64-72.

Laudon, K. C. (1996) Markets and Privacy. *Communications of the ACM,* Vol. 39, No. 9, pp. 92-104.

Liu, C. N., and Zhu, X. W. (2009) A Study on CRM Technology Implementation and Application Practices. In: *International Conference on Computational Intelligence and Natural Computing.* Wuhan, June 2009. IEEE Computer Society, pp. 367-370.

Madria, S. K., Bhowmick, S. S., Ng, W. K., and Lim, E. P. (1999) Research Issues in Web Data Mining. In Proceedings of: *Data Warehousing and Knowledge Discovery, First International Conference, DaWaK '99,* pp. 303–312.

Magnini, V. P., Honeycutt Jr., E. D., and Hodge, S. K. (2003) Data Mining for Hotel Firms: Use and Limitations. *Cornell Hotel and Restaurant Administration Quarterly 44,* No. 2, pp. 94-105.

McCarty, J. A., and Hastak, M. (2007) Segmentation Approaches in Data-Mining: A Comparison of RFM, CHAID, and Logistic Regression. *Journal of Business Research,* Vol. 60, No. 6, pp. 656-652.

Mobasher, B., Dai, H., Luo, T., and Nakagawa, M. (2002) Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization. *Data Mining and Knowledge Discovery,* Vol. 6, 61-82.

Mobasher, B., Jain, N., Han, E. H., and Srivastava, J. (1996) Web Mining: Pattern Discovery from World wide Web Transactions. Technical Report TR 96-050, University of Minnesota, Dept. of Computer Science, Minneapolis.

Nasraoui, O., Soliman, M., Saka, E., Badia, A., and Germain, R. (2008) A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites. *IEEE Transactions on Knowledge and Data Engineering,* Vol. 20, No. 2, pp. 202-215.

Ngai, E. W. T., Xiu, L., and Chau, D. C. K. (2009) Application of Data Mining Techniques in Customer Relationship Management: A literature review and classification. *Expert Systems with Applications,* Vol. 36, pp. 2592-2602.

Olmeda, I., and Sheldon, P. J. (2001) Data Mining Techniques and Applications for Tourism Internet Marketing. *Journal of Travel and Tourism Marketing.* Vol. 11, pp. 1-20.

OECD (1980) Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, available from: http://www.oecd.org/document/18/0,3746,en_2649_34255_1815186_1_1_1_1,00.html (Accessed 10/12/2011).

Parvatiyar, A., and Sheth, J. N. (2001) Customer Relationship Management: Emerging Practice, Process, and Discipline. *Journal of Economic and Social Research,* Vol. 3, No. 2, pp. 1-34.

Payne, A. (2002) The value creation process in customer relationship management. *Insight Interactive,* pp. 1-17.

Purandare, P. (2008) Web Mining: A Key to Improve Business on Web. In: Ajith, P. Abraham. *Proceedings of the IADIS European Conference on Data Mining, Amsterdam.* The Netherlands, July 24-26. IADIS, pp. 155-159.

Pyo, S., Uysal, M., and Chang, H. (2002) Knowledge Discovery in Database for Tourist Destinations. *Journal of Travel Research,* Vol. 40, pp. 396-403.

Roddick, J. F. and Fule, P. (2003) A System for Detecting Privacy and Ethical Sensitivity in Data Mining Results. Technical Report SIE-03-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia.

Rygielski, C., Wang, J. C., and Yen, D. C. (2002) Data Mining Techniques for Customer Relationship Management. *Technology In Society,* Vol. 24, pp. 483-502.

Shaw, M. J., Subramaniam, C., Tan, G. W., and Welge, M. E. (2001) Knowledge Management and Data Mining for Marketing. *Decision Support Systems, Vol.* 31, pp. 127-137.

Sohrabi, B., and Khanlari, A. (2007) Customer Lifetime Value Measurement Based on RFM Model. *Iranian Accounting & Auditing Review,* Vol. 14, No. 47, pp. 7-20.

Srivastava, J., Cooley, R. W., Deshpande, M., Tan, P. N. (2000), Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations,* Vol. 1 No.2, pp. 1-12.

Tavani, H. T. (1999) Informational Privacy, Data Mining, and the Internet. *Ethics and Information Technology,* Vol. 1. pp. 137-145.

Thanuja, V., Venkateswarlu, B., and Anjanjeyulu, G. S. G. N. (2011) Applications of Data Mining in Customer Relationship Management. *Journal of Computer and Mathematical Sciences,* Vol. 2, No. 3, pp. 423-433.

Thearling, K. (2010) Data Mining for CRM. In: Maimon, O. and Rokach, L. Data Mining and Knowledge Discovery Handbook, 2nd ed. LLC, Springer Science & Business Media, pp. 1181-1188.

Thearling, K. (2000) Data Mining and Customer Relationships. Available from: http://www.thearling.com/text/whexcerpt/whexcerpt.htm .

Viswanathan, M. (2007) Distributed Data Mining in a Ubiquitous Healthcare Framework. *Advanves in Artificial Intelligence, Lecture Notes in Computer Science,* Vol. 4509, pp. 261-271.

Watson, R. T., G. M. Zinkhan and L. F. Pitt (2000). Integrated Internet Marketing. *Communications of the ACM*, Vol. 43, No. 6, pp. 97-102.

Wahlstrom, K., Roddick, J. F., Sarre, R., Estivill-Castro, V. and de Vries, D. (2002), On the Ehics of Data Mining, *Technical Report KDM-02-006*, Flinders University of South Australia.

Weiss, G. M. (2005) Data Mining in Telecommunications. In: Maimon, O., and Rokach, L. (eds.), Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, Kluwer, pp. 1189-1201.

Wilson, R. F. (2000) Preparing a Customer Profile for Your Internet Marketing Plan, *Web Marketing Today*, April 1, 2000. Available from: http://www.wilsonweb.com/wmt5/customers-profile.htm .

Wong, J. Y., Chen, H. J., Chung, P. H., and Kao, N. C. (2006) Identifying Valuable Travelers and Their Next Foreign Destination by the Application of Data Mining Techniques. *Asia Pacific Journal of Tourism Research*, Vol. 11, No. 4, pp. 355-373.

Wu, X., Yu, P. S., Piatetsky-Shapiro, G., Cercone, N., Lin, T. Y., Kotagiri, R., and Wah, B. W. (2003) Data Mining: How Research Meets Practical Development? *Knowledge and Information Systems*, Vol. 5, No. 2, pp. 248-261.

Wyner. G. A. (1996) Customer Valuation: Linking Behavior and Economics. *Marketing Research.* Vol. 8, No. 2, pp. 36-38.

Xu, M., and Walton, J. (2005) Gaining Customer Knowledge through Analytical CRM. *Industrial Management and Data Systems,* Vol. 105, No. 7, pp. 955-971.

Yang, A. X. (2004) How to Develop New Approaches to RFM Segmentation. *Journal of Targeting, Measurement and Analysis for Marketing,*Vol. 13, No. 1, pp. 50-60.

Yina, W. (2010) Application of Customer Relationship Management in Health Care. In*: Proceedings of Second International Conference on MultiMedia and Information Technology*. Kaifeng, April 2010, pp. 52-55.